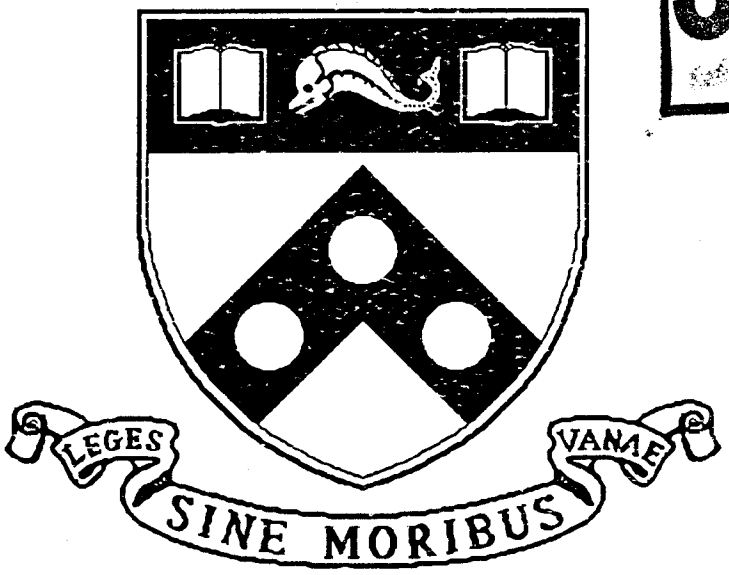
What's So Special About Kruskal's Theorem And The Ordinal Γ_0? A Survey Of Some Results In Proof Theory

MS-CIS-93-82
LOGIC & COMPUTATION 72

Jean H. Gallier



University of Pennsylvania
School of Engineering and Applied Science
Computer and Information Science Department
Philadelphia, PA 19104-6389

September 1993

MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
		technical report

4. TITLE AND SUBTITLE

What's So Special About Kruskal's Theorem and The Ordinal Γ_0 ? A Survey of Some Results In Proof Theory

5. FUNDING NUMBERS

DAAL03-89-C-0031

6. AUTHOR(S)

Jean H. Gallier

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Pennsylvania
Department of Computer and Information Sciences
200 S. 33rd Street
Philadelphia, PA 19104-6389

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ARO 26779.26-MA-AI

11. SUPPLEMENTARY NOTES

The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This paper consists primarily of a survey of results of Harvey Friedman about some proof theoretic aspects of various forms of Krusal's tree theorem, and in particular the connection with the ordinal Γ_0. We also include a fairly extensive treatment of normal functions on the countable ordinals, and we give a glimpse of Veblen Hierarchies, some subsystems of second-order logic, slow-growing and fast-growing hierarchies including Girard's result, and Goodstein sequences. The central theme of this paper is a powerful theorem due to Kruskal, the "tree theorem", as well as a "finite miniaturization" of Kruskal's theorem due to Harvey Friedman. These versions of Kruskal's theorem are remarkable from a proof-theoretic point of view because they are *not* provable in relatively strong logical systems. They are examples of so-called "natural independence phenomena", which are considered by more logicians as more natural than the mathematical incompleteness results first discovered by Gödel. Kruskal's tree theorem also plays a fundamental role in computer science, because it is one of the main tools for showing that certain orderings on trees are well founded. These orderings play a crucial role in proving the termination of systems of rewrite rules and the correctness of Knuth-Bandix completion procedures. There is also a close connection between a certain infinite countable ordinal called Γ_0 and Kruskal's theorem. Previous definitions of the function involved in this connection are known to be incorrect, in that, the function is not monotonic. We offer a repaired definition of this function, and explore briefly the consequences of its existence.

14. SUBJECT TERMS			15. NUMBER OF PAGES
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

WHAT'S SO SPECIAL ABOUT KRUSKAL'S THEOREM AND THE ORDINAL Γ_0? A SURVEY OF SOME RESULTS IN PROOF THEORY

Jean H. Gallier
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, Pa 19104

September 30, 1993

This research was partially supported by the National Science Foundation under Grant No. DCR-86-07156, and ONR under Grant No. N00014-88-K-0593.

WHAT'S SO SPECIAL ABOUT KRUSKAL'S THEOREM AND THE ORDINAL Γ_0? A SURVEY OF SOME RESULTS IN PROOF THEORY

Jean H. Gallier

Abstract: This paper consists primarily of a survey of results of Harvey Friedman about some proof theoretic aspects of various forms of Kruskal's tree theorem, and in particular the connection with the ordinal Γ_0. We also include a fairly extensive treatment of normal functions on the countable ordinals, and we give a glimpse of Veblen hierarchies, some subsystems of second-order logic, slow-growing and fast-growing hierarchies including Girard's result, and Goodstein sequences. The central theme of this paper is a powerful theorem due to Kruskal, the "tree theorem", as well as a "finite miniaturization" of Kruskal's theorem due to Harvey Friedman. These versions of Kruskal's theorem are remarkable from a proof-theoretic point of view because they are *not* provable in relatively strong logical systems. They are examples of so-called "natural independence phenomena", which are considered by most logicians as more natural than the metamathematical incompleteness results first discovered by Gödel. Kruskal's tree theorem also plays a fundamental role in computer science, because it is one of the main tools for showing that certain orderings on trees are well founded. These orderings play a crucial role in proving the termination of systems of rewrite rules and the correctness of Knuth-Bendix completion procedures. There is also a close connection between a certain infinite countable ordinal called Γ_0 and Kruskal's theorem. Previous definitions of the function involved in this connection are known to be incorrect, in that, the function is not monotonic. We offer a repaired definition of this function, and explore briefly the consequences of its existence.

1 Introduction

This paper consists primarily of a survey of results of Harvey Friedman [47] about some proof theoretic aspects of various forms of Kruskal's tree theorem [28], and in particular the connection with the ordinal Γ_0. Initially, our intention was to restrict ourselves to Kruskal's tree theorem and Γ_0. However, as we were trying to make this paper as self contained as possible, we found that it was necessary to include a fairly extensive treatment of normal functions on the countable ordinals. Thus, we also give a glimpse of Veblen hierarchies, some subsystems of second-order logic, slow-growing and fast-growing hierarchies including Girard's result, and Goodstein sequences.

The central theme of this paper is a powerful theorem due to Kruskal, the "tree theorem", as well as a "finite miniaturization" of Kruskal's theorem due to Harvey Friedman. These versions of Kruskal's theorem are remarkable from a proof-theoretic point of view because they are *not* provable in relatively strong logical systems. They are examples of so-called "natural independence phenomena", which are considered by most logicians as more natural than the metamathematical incompleteness results first discovered by Gödel.

Kruskal's tree theorem also plays a fundamental role in computer science, because it is one of the main tools for showing that certain orderings on trees are well founded. These orderings play a crucial role in proving the termination of systems of rewrite rules and the correctness of Knuth-Bendix completion procedures [27].

There is also a close connection between a certain infinite countable ordinal called Γ_0 (Feferman [13], Schütte [46]) and Kruskal's theorem. This connection lies in the fact that there is a close relationship between the embedding relation $\preceq$ on the set T of finite trees (see definition 4.11) and the well-ordering $\leq$ on the set $\mathcal{O}(\Gamma_0)$ of all ordinals $< \Gamma_0$. Indeed, it is possible to define a function $h : T \to \mathcal{O}(\Gamma_0)$ such that h is (1). surjective, and (2). preserves order, that is, if $s \preceq t$, then $h(s) \leq h(t)$. Previous definitions of this function are known to be incorrect, in that, the function is not monotonic. We offer a repaired definition of this function, and explore briefly the consequences of its existence.

We believe that there is a definite value in bringing together a variety of topics revolving around a common theme, in this case, ordinal notations and their use in mathematical logic. We are hoping that our survey will help in making some beautiful but seemingly rather arcane tools and techniques known to more researchers in logic and theoretical computer science.

The paper is organized as follows. Section 2 contains all the definitions about pre-orders, well-founded orderings, and well-quasi orders (WQO's), needed in the rest of the paper. Higman's theorem for WQO's on strings is presented in section 3. Several versions

of Kruskal's tree theorem are presented in section 4. Section 5 is devoted to several versions of the finite miniaturization of Kruskal's theorem due to Harvey Friedman. Section 6 is a fairly lengthy presentation of basic facts about the countable ordinals, normal functions, and Γ_0. Most of this material is taken from Schütte [46], and we can only claim to have presented it our own way, and hopefully made it more accessible. Section 7 gives a glimpse at Veblen hierarchies. A constructive system of notations for Γ_0 is presented in section 8. The connection between Kruskal's tree theorem and Γ_0 due to Friedman is presented in section 9. A brief discussion of some relevant subsystems of second-order arithmetic occurs in section 10. An introduction to the theory of term orderings is presented in section 11, including the recursive path ordering and the lexicographic path ordering. A glimpse at slow-growing and fast-growing hierarchies is given in section 12. Finally, constructive proofs of Higman's lemma are briefly discussed in section 13.

2 Well Quasi-Orders (WQO's)

We let $\mathbf{N}$ denote the set $\{0, 1, 2, \ldots\}$ of natural numbers, and $\mathbf{N}_+$ denote the set $\{1, 2, \ldots\}$ of positive natural numbers. Given any $n \in \mathbf{N}_+$, we let $[n]$ denote the finite set $\{1, 2, \ldots, n\}$, and we let $[0] = \emptyset$. Given a set S, a *finite sequence* u over S, or *string* over S, is a function $u : [n] \to S$, for some $n \in \mathbf{N}$. The integer n is called the *length* of u and is denoted by $|u|$. The special sequence with domain $\emptyset$ is called the *empty sequence*, or *empty string*, and will be denoted by e. Strings can be *concatenated* in the usual way: Given two strings $u : [m] \to S$ and $v : [n] \to S$, their *concatenation* denoted by $u.v$ or uv, is the string $uv : [m + n] \to S$ such that, $uv(i) = u(i)$ if $1 \leq i \leq m$, and $uv(i) = v(i - m)$ if $m + 1 \leq i \leq m + n$. Clearly, concatenation is associative and e is an identity element. Occasionally, a finite sequence u of length n will be denoted as $\langle u_1, \ldots, u_n \rangle$ (denoting $u(i)$ as u_i), or as $u_1 \ldots u_n$. Strings of length 1 are identified with elements of S. The set of all strings over S is denoted as S^*.

An *infinite sequence* is a function $s : \mathbf{N}_+ \to S$. An infinite sequence s is also denoted by $(s_i)_{i \geq 1}$, or by $\langle s_1, s_2, \ldots, s_i, \ldots \rangle$. Given an infinite sequence $s = (s_i)_{i \geq 1}$, an *infinite subsequence* of s is any infinite sequence $s' = (s'_j)_{j \geq 1}$ such that there is a strictly monotonic function[1] $f : \mathbf{N}_+ \to \mathbf{N}_+$, and $s'_i = s_{f(i)}$ for all $i > 0$. An infinite subsequence s' of s associated with the function f is also denoted as $s' = (s_{f(i)})_{i \geq 1}$.

We now review preorders and well-foundedness.

Definition 2.1 Given a set A, a binary relation $\preceq \subseteq A \times A$ on the set A is a *preorder*

[1] A function $f : \mathbf{N}_+ \to \mathbf{N}_+$ is *strictly monotonic* (or *increasing*) iff for all $i, j > 0$, $i < j$ implies that $f(i) < f(j)$.

(or *quasi-order*) iff it is reflexive and transitive. A preorder that is also antisymmetric is called a *partial order*. A preorder is *total* iff for every $x, y \in A$, either $x \preceq y$ or $y \preceq x$. The relation $\succeq$ is defined such that $x \succeq y$ iff $y \preceq x$, the relation $\prec$ such that

$$x \prec y \quad \text{iff} \quad x \preceq y \quad \text{and} \quad y \not\preceq x,$$

the relation $\succ$ such that $x \succ y$ iff $y \prec x$, and the equivalence relation $\approx$ such that

$$x \approx y \quad \text{iff} \quad x \preceq y \quad \text{and} \quad y \preceq x.$$

We say that x and y are *incomparable* iff $x \not\preceq y$ and $y \not\preceq x$, and this is also denoted by $x \mid y$.

Given two preorders $\preceq_1$ and $\preceq_2$ on a set A, $\preceq_2$ is an *extension* of $\preceq_1$ iff $\preceq_1 \subseteq \preceq_2$. In this case, we also say that $\preceq_1$ is a *restriction* of $\preceq_2$.

Definition 2.2 Given a preorder $\preceq$ over a set A, an infinite sequence $(x_i)_{i \geq 1}$ is an *infinite decreasing chain* iff $x_i \succ x_{i+1}$ for all $i \geq 1$. An infinite sequence $(x_i)_{i \geq 1}$ is an *infinite antichain* iff $x_i \mid x_j$ for all i, j, $1 \leq i < j$. We say that $\preceq$ is *well-founded* and that $\succ$ is *Noetherian* iff there are no infinite decreasing chains w.r.t. $\succ$.

We now turn to the fundamental concept of a well quasi-order. This concept goes back at least to Janet [23], whose paper appeared in 1920, as recently noted by Pierre Lescanne [31]. Irving Kaplanski also told me that this concept is defined and used in his Ph.D thesis [25] (1941). The concept was further investigated by Higman [22], Kruskal [28], and Nash-Williams [36], among the forerunners.

Definition 2.3 Given a preorder $\preceq$ over a set A, an infinite sequence $(a_i)_{i \geq 1}$ of elements in A is termed *good* iff there exist positive integers i, j such that $i < j$ and $a_i \preceq a_j$, and otherwise, it is termed a *bad* sequence. A preorder $\preceq$ is a *well quasi-order*, abbreviated as *wqo*, iff every infinite sequence of elements of A is good.

Among the various characterizations of *wqo*'s, the following ones are particularly useful.

Lemma 2.4 Given a preorder $\preceq$ on a set A, the following conditions are equivalent:

1. Every infinite sequence is good (w.r.t. $\preceq$).
2. There are no infinite decreasing chains and no infinite antichains (w.r.t. $\preceq$).
3. Every preorder extending $\preceq$ (including $\preceq$ itself) is well-founded.

Proof. (1) $\Longrightarrow$ (2). Suppose that $(x_i)_{i\geq 1}$ is an infinite sequence over A such that $x_i \succ x_{i+1}$ for all $i \geq 1$. Hence, for every $i \geq 1$,

$$x_{i+1} \preceq x_i, \quad \text{and} \quad x_i \not\preceq x_{i+1}. \qquad (*)$$

Since $\preceq$ satisfies (1), there exist some integers $i, j > 0$ such that $i < j$ and $x_i \preceq x_j$. If $j = i+1$, this contradicts $(*)$. If $j > (i+1)$, by transitivity of $\preceq$, since $x_{j-1} \preceq \ldots \preceq x_i \preceq x_j$, we have $x_{j-1} \preceq x_j$, contradicting $(*)$. Hence there are no infinite decreasing sequences, that is, $\preceq$ is well-founded. Also, it is clear that the existence of an infinite antichain would contradict (1).

(2) $\Longrightarrow$ (3). This proof is identical to the first part of the proof of (1) $\Longrightarrow$ (2).

(3) $\Longrightarrow$ (1). If (1) fails, then there is some infinite sequence $s = (x_i)_{i\geq 1}$ such that $x_i \not\preceq x_j$ for all i, j, $1 \leq i < j$. But then, we can extend $\preceq$ to a preorder $\preceq'$ such that s becomes an infinite decreasing chain in $\preceq'$, contradicting (3). □

It is interesting to observe that the property of being a *wqo* is substantially stronger that being well-founded. Indeed, it is not true in general that any preorder extending a given well-founded preorder is well-founded. However, by (3) of lemma 2.4, this property characterizes a *wqo*. Every preorder on a finite set (including the equality relation) is a *wqo*, and by (3) of lemma 2.4, every partial ordering that is total and well-founded is a *wqo* (such orderings are called *well-orderings*).

The following lemma turns out to be the key to the proof of Kruskal's theorem. It is implicit in Nash-Williams [36], lemma 1, page 833.

Lemma 2.5 Given a preorder $\preceq$ on a set A, the following are equivalent:

(1) $\preceq$ is a *wqo* on A.

(2) Every infinite sequence $s = (s_i)_{i\geq 1}$ over A contains some infinite subsequence $s' = (s_{f(i)})_{i\geq 1}$ such that $s_{f(i)} \preceq s_{f(i+1)}$ for all $i > 0$.

Proof. It is clear that (2) implies (1). Next, assume that $\preceq$ is a *wqo*. We say that a member s_i of a sequence s is *terminal* iff there is no $j > i$ such that $s_i \preceq s_j$. We claim that the number of terminal elements in the sequence s is finite. Otherwise, the infinite sequence t of terminal elements in s is a bad sequence (because if the sequence t was good, then we would have $s_h \preceq s_k$ for two terminal elements in s, contradicting the fact that s_h is terminal), and this contradicts the fact that $\preceq$ is a *wqo*. Hence, there is some $N > 0$ such that s_i is not terminal for every $i \geq N$. We can define a strictly monotonic function f inductively as follows. Let $f(1) = N$, and for any $i \geq 1$, let $f(i+1)$ be the least integer such that

$s_{f(i)} \preceq s_{f(i+1)}$ and $f(i+1) > f(i)$ (since every element $s_{f(i)}$ is not terminal by the choice of N and the definition of f, such an element exists). The infinite subsequence $s' = (s_{f(i)})_{i \geq 1}$ has the property stated in (2). □

As a corollary of lemma 2.5, we obtain another result of Nash-Williams [36]. Given two preorders $\langle \preceq_1, A_1 \rangle$ and $\langle \preceq_2, A_2 \rangle$, the cartesian product $A_1 \times A_2$ is equipped with the preorder $\preceq$ defined such that $(a_1, a_2) \preceq (a_1', a_2')$ iff $a_1 \preceq_1 a_1'$ and $a_2 \preceq_2 a_2'$.

Lemma 2.6 If $\preceq_1$ and $\preceq_2$ are *wqo*, then $\preceq$ is a *wqo* on $A_1 \times A_2$.

Proof. Consider any infinite sequence s in $A_1 \times A_2$. This sequence is formed of pairs $(s_i', s_i'') \in A_1 \times A_2$, and defines an infinite sequence $s' = (s_i')_{i \geq 1}$ over A_1 and an infinite sequence $s'' = (s_i'')_{i \geq 1}$ over A_2. By lemma 2.5, since $\preceq_1$ is a *wqo*, there is some infinite subsequence $t' = (s'_{f(i)})_{i \geq 1}$ of s' such that $s'_{f(i)} \preceq_1 s'_{f(i+1)}$ for all $i > 0$. Since $\preceq_2$ is also a *wqo* and $t'' = (s''_{f(i)})_{i \geq 1}$ is an infinite sequence over A_2, there exist some i, j such that $f(i) < f(j)$ and $s''_{f(i)} \preceq_2 s''_{f(j)}$. Then, we have $(s'_{f(i)}, s''_{f(i)}) \preceq (s'_{f(j)}, s''_{f(j)})$, which shows that the sequence s is good, and that $\preceq$ is a *wqo*. □

In turn, lemma 2.6 yields an interesting result due to Dickson [12], published in 1913!

Lemma 2.7 Let n be any integer such that $n > 1$. Given any infinite sequence $(s_i)_{i \geq 1}$ of n-tuples of natural numbers, there exist positive integers i, j such that $i < j$ and $s_i \preceq_n s_j$, where $\preceq_n$ is the partial order on n-tuples of natural numbers induced by the natural ordering $\leq$ on $\mathbf{N}$.

Proof. The proof follows immediately by observing that $\leq$ is a *wqo* on $\mathbf{N}$ and that lemma 2.6 extends to any $n > 1$ by a trivial induction. □

Next, given a *wqo* $\preceq$ on a set A, we shall extend $\preceq$ to the set of strings A^*, and prove what is known as *Higman's theorem* [22].

3 WQO's On Strings, Higman's Theorem

Our presentation of Higman's theorem is inspired by Nash-Williams's proof of a similar theorem ([36], lemma 2, page 834), and is also very similar to the proof given by Steve Simpson ([47], lemma 1.6, page 92). Nash-Williams's proof is not entirely transparent, and Simpson's proof appeals to Ramsey's theorem. Using lemma 2.5, it is possible to simplify the proof. A proof along this line has also been given by Jean Jacques Levy in some unpublished notes [33] that came mysteriously in my possession.

Definition 3.1 Let $\sqsubseteq$ be a preorder on a set A. We define the preorder $\ll$ (*string embedding*) on A^* as follows: $e \ll u$ for each $u \in A^*$, and, for any two strings $u = u_1u_2 \ldots u_m$ and $v = v_1u_2 \ldots v_n$, $1 \leq m \leq n$,

$$u_1u_2 \ldots u_m \ll v_1v_2 \ldots v_n$$

iff there exist integers $j_1, \ldots, j_m$ such that $1 \leq j_1 < j_2 < \ldots < j_{m-1} < j_m \leq n$ and

$$u_1 \sqsubseteq v_{j_1}, \ \ldots, \ u_m \sqsubseteq v_{j_m}.$$

It is easy to show that $\ll$ is a preorder, and we leave as an exercise to show that $\ll$ is a partial order if $\sqsubseteq$ is a partial order. It is also easy to check that $\ll$ is the least preorder on A^* satisfying the following two properties:

(1) (deletion property) $uv \ll uav$, for all $u, v \in A^*$ and $a \in A$;

(2) (monotonicity) $uav \ll ubv$ whenever $a \sqsubseteq b$, for all $u, v \in A^*$ and $a, b \in A$.

Theorem 3.2 (Higman) If $\sqsubseteq$ is a *wqo* on A, then $\ll$ is a *wqo* on A^*.

Proof. Assume that $\ll$ is not a *wqo* on A^*. Then, there is at least one bad sequence from A^*. Following Nash-Williams, we define a *minimal bad sequence* t inductively as follows. Let t_1 be a string of minimal length starting a bad sequence. If $t_1, \ldots, t_n$ have been defined, let t_{n+1} be a string of minimal length such that there is a bad sequence whose first n elements are $t_1, \ldots, t_n$. Note that we must have $|t_i| \geq 1$ for all $i \geq 1$, since otherwise the sequence t is not bad (since $e \ll u$ for each $u \in A^*$). Since $|t_i| \geq 1$ for all $i \geq 1$, let

$$t_i = a_is_i,$$

where $a_i \in A$ is the leftmost symbol in t_i. The elements a_i define an infinite sequence $a = (a_i)_{i \geq 1}$ in A, and the s_i define an infinite sequence $s = (s_i)_{i \geq 1}$ in A^*. Since $\sqsubseteq$ is a *wqo* on A, by lemma 2.5, there is an infinite subsequence $a' = (a_{f(i)})_{i \geq 1}$ of a such that $a_{f(i)} \sqsubseteq a_{f(i+1)}$ for all $i > 0$. We claim that the infinite subsequence $s' = (s_{f(i)})_{i \geq 1}$ of s is good. Otherwise, if $s' = (s_{f(i)})_{i \geq 1}$ is bad, there are two cases.

Case 1: $f(1) = 1$. Then, the infinite sequence $s' = (s_{f(i)})_{i \geq 1}$ is a bad sequence with $|s_1| < |t_1|$, contradicting the minimality of t.

Case 2: $f(1) > 1$. Then, the infinite sequence

$$t' = \langle t_1, \ldots, t_{f(1)-1}, s_{f(1)}, s_{f(2)}, \ldots, s_{f(j)}, \ldots \rangle$$

is also bad, because $t_k = a_ks_k$ for all $k \geq 1$ and $t_i \ll s_{f(j)}$ implies that $t_i \ll t_{f(j)}$ by the definition of $\ll$. But $|s_{f(1)}| < |t_{f(1)}|$, and this contradicts the minimality of t.

Since the sequence $s' = (s_{f(i)})_{i \geq 1}$ is good, there are some positive integers i, j such that $f(i) < f(j)$ and $s_{f(i)} \ll s_{f(j)}$. Since the infinite sequence $a' = (a_{f(i)})_{i \geq 1}$ was chosen such that $a_{f(i)} \sqsubseteq a_{f(i+1)}$ for all $i > 0$, by the definition of $\ll$, we have

$$a_{f(i)} s_{f(i)} \ll a_{f(j)} s_{f(j)},$$

that is, $t_{f(i)} \ll t_{f(j)}$ (since $t_k = a_k s_k$ for all $k \geq 1$). But this shows that the sequence t is good, contradicting the initial assumption that t is bad. □

A theorem similar to theorem 3.2 applying to finite subsets of A can be shown. Following Nash-Williams [36], let $\mathcal{F}(S)$ denote the set of all finite subsets of S. Given any two subsets A, B of S, a function $f : A \to B$ is *non-descending* if $a \sqsubseteq f(a)$ for every $a \in A$. The set $\mathcal{F}(S)$ is equipped with the preorder $\ll$ defined as follows: $\emptyset \ll A$ for every $A \in \mathcal{F}(S)$, and for any two nonempty subsets $A, B \in \mathcal{F}(S)$, $A \ll B$ iff there is an injective non-descending function $f : A \to B$. The proof of theorem 3.2 can be trivially modified to obtain the following.

Theorem 3.3 (Nash-Williams) If $\sqsubseteq$ is a *wqo* on A, then $\ll$ is a *wqo* on $\mathcal{F}(A)$.

We now turn to trees.

4 WQO's On Trees, Kruskal's Tree Theorem

First, we review the definition of trees in terms of tree domains.

Definition 4.1 A *tree domain* D is a nonempty subset of strings in $\mathbf{N}_+^*$ satisfying the conditions:

(1) For all $u, v \in \mathbf{N}_+^*$, if $uv \in D$ then $u \in D$.

(2) For all $u \in \mathbf{N}_+^*$, for every $i \in \mathbf{N}_+$, if $ui \in D$ then, for every j, $1 \leq j \leq i$, $uj \in D$.

The elements of D are called *tree addresses* or *nodes*. We now consider labeled trees.

Definition 4.2 Given any set Σ of labels, a Σ-*tree (or term)* is any function $t : D \to \Sigma$, where D is a tree domain denoted by $dom(t)$.

Hence, a labeled tree is defined by a tree domain D and a labeling function t with domain D and range Σ. The tree address e is called the *root* of t, and its label $t(e)$ is denoted as $root(t)$. A tree is *finite* iff its domain is finite. In the rest of this paper, only finite trees will be considered. The set of all finite Σ-trees is denoted as T_Σ.

Definition 4.3 Given a (finite) tree t, the number of tree addresses in $dom(t)$ is denoted by $|t|$. The depth of a tree t is defined as $depth(t) = max(\{|u| \mid u \in dom(t)\})$. The number of immediate successors of the root of a tree is denoted by $rank(t)$, and it is defined formally as the number of elements in the set $\{i \mid i \in \mathbf{N}_+ \text{ and } i \in dom(t)\}$. Given a tree t and some tree address $u \in dom(t)$, the *subtree of t rooted at u* is the tree t/u whose domain is the set $\{v \mid uv \in dom(t)\}$ and such that $t/u(v) = t(uv)$ for all v in $dom(t/u)$.

A tree t such that $rank(t) = 0$ is a one-node tree, and if $root(t) = f$, t will also be denoted by f. Given any $k \geq 1$ trees $t_1, \ldots, t_k$ and any element $f \in \Sigma$, the tree $t = f(t_1, \ldots, t_k)$ is the tree whose domain is the set

$$\{\epsilon\} \cup \bigcup_{i=1}^{i=k} \{iu \mid u \in dom(t_i)\},$$

and whose labeling function is defined such that $t(\epsilon) = f$ and $t(iu) = t_i(u)$, for $u \in dom(t_i)$, $1 \leq i \leq k$. It is well known that every finite tree t is either a one-node tree, or can be written uniquely as $t = f(t/1, \ldots, t/k)$, where $f = root(\epsilon)$, and $k = rank(t)$. It is also convenient to introduce the following abbreviations. Let $\sqsubseteq$ be a binary relation on trees. Then

$$s \sqsubseteq f(\ldots, s, \ldots)$$

is an abbreviation for $s \sqsubseteq f(s_1, \ldots, s_{i-1}, s, s_{i+1}, \ldots, s_n)$,

$$f(\ldots) \sqsubseteq f(\ldots, s, \ldots)$$

is an abbreviation for $f(s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n) \sqsubseteq f(s_1, \ldots, s_{i-1}, s, s_{i+1}, \ldots, s_n)$,

$$f(\ldots, s, \ldots) \sqsubseteq g(\ldots, t, \ldots)$$

is an abbreviation for $f(s_1, \ldots, s_{i-1}, s, s_{i+1}, \ldots, s_n) \sqsubseteq g(s_1, \ldots, s_{i-1}, t, s_{i+1}, \ldots, s_n)$, for some trees s, t, $s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n$, $1 \leq i \leq n$. When $n = 1$, these are understood as $s \sqsubseteq f(s)$, $f \sqsubseteq f(s)$, and $f(s) \sqsubseteq g(t)$.

4.1 Kruskal's Theorem, Version 1

Assuming that Σ is preordered by $\sqsubseteq$, we define a preorder $\preceq$ on Σ-trees extending $\sqsubseteq$ in the following way.

Definition 4.4 Assume that $\sqsubseteq$ is a preorder on Σ. The preorder $\preceq$ on T_Σ (*homeomorphic embedding*) is defined inductively as follows: Either

(1) $f \preceq g(t_1, \ldots, t_n)$ iff $f \sqsubseteq g$; or

(2) $s \preceq g(\ldots, t, \ldots)$ iff $s \preceq t$; or

(3) $f(s_1, \ldots, s_m) \preceq g(t_1, \ldots, t_n)$ iff $f \sqsubseteq g$, and there exist some integers $j_1, \ldots, j_m$ such that $1 \leq j_1 < j_2 < \ldots < j_{m-1} < j_m \leq n$, $1 \leq m \leq n$, and

$$s_1 \preceq t_{j_1}, \ \ldots, \ s_m \preceq t_{j_m}.$$

Note that (1) can be viewed as the special case of (3) for which $m = 0$, and $n = 0$ is possible. It is easy to show that $\preceq$ is a preorder. One can also show that $\preceq$ is a partial order if $\sqsubseteq$ is a partial order. This can be shown by observing that $s \preceq t$ implies that $depth(s) \leq depth(t)$. Hence, if $s \preceq t$ and $t \preceq s$, we have $depth(s) = depth(t)$ and $rank(s) = rank(t)$ (since only case (1) or (3) can apply). Then, we can show that $s = t$ by induction on the depth of trees.

It is also easy to show that the preorder $\preceq$ can be defined as the least preorder satisfying the following properties:

(1) $s \preceq f(\ldots, s, \ldots)$;

(2) $f(\ldots) \preceq f(\ldots, s, \ldots)$;

(3) $f(\ldots, s, \ldots) \preceq g(\ldots, t, \ldots)$ whenever $f \sqsubseteq g$ and $s \preceq t$.

We now prove a version of Kruskal's theorem [28].

Theorem 4.5 (Kruskal's tree theorem) If $\sqsubseteq$ is a *wqo* on Σ, then $\preceq$ is a *wqo* on T_Σ.

Proof. Assume that $\preceq$ is not a *wqo* on T_Σ. As in the proof of theorem 3.2, we define a minimal bad sequence t of elements of T_Σ satisfying the following properties:

(i) $|t_1| \leq |t'_1|$ for all bad sequences t';

(ii) $|t_{n+1}| \leq |t'_{n+1}|$ for all bad sequences t' such that $t'_i = t_i$, $1 \leq i \leq n$.

We claim that $|t_i| \geq 2$ for all but finitely many $i \geq 1$. Otherwise, the sequence of one-node trees in t must be infinite, and since $\sqsubseteq$ is a *wqo*, by clause (1) of the definition of $\preceq$, there are $i, j > 0$ such that $i < j$ and $t_i \preceq t_j$, contradicting the fact that t is bad.

Let $s = (s_i)_{i \geq 1}$ be the infinite subsequence of t consisting of all trees having at least two nodes, and let $f = (f_i)_{i \geq 1}$ be the infinite sequence over Σ defined such that $f_i = root(s_i)$ for every $i \geq 1$. Since $\sqsubseteq$ is a *wqo* over Σ, by lemma 2.5, there is some infinite subsequence $f' = (f_{\varphi(i)})_{i \geq 1}$ of f such that $f_{\varphi(i)} \sqsubseteq f_{\varphi(i+1)}$ for all $i \geq 1$. Let

$$\mathcal{D} = \{s_{\varphi(i)}/j \mid i \geq 1, \ 1 \leq j \leq rank(s_{\varphi(i)})\}.$$

We claim that $\preceq$ is a *wqo* on $\mathcal{D}$. Otherwise, let $r = \langle r_1, r_2, \ldots, r_j, \ldots \rangle$ be a bad sequence in $\mathcal{D}$. Because r is bad, it contains a bad subsequence $r' = \langle r'_1, r'_2, \ldots, r'_j, \ldots \rangle$ with the following property: if $i < j$, then r'_i is a subtree of a tree t_p and r'_j is a subtree of a tree t_q such that $p < q$. Indeed, every t_i only has finitely many subtrees, and r being bad must contain an infinite number of distinct trees. Thus, we consider a bad sequence r with the additional property that if $i < j$, then r_i is a subtree of a tree t_p and r_j is a subtree of a tree t_q such that $p < q$. Let n be the index of the first tree in the sequence t such that $t_n/j = r_1$ for some j. If $n = 1$, since $|r_1| < |t_1|$ and the sequence r is bad, this contradicts the fact that t is a minimal bad sequence. If $n > 1$, then the sequence

$$\langle t_1, t_2, \ldots, t_{n-1}, r_1, r_2, \ldots, r_j, \ldots \rangle$$

is bad, since by clause (ii) of the definition of $\preceq$, for any k s.t. $1 \leq k \leq n-1$, $t_k \preceq r_j$ would imply that $t_k \preceq t_h$ for some t_h and l such that $r_j = t_h/l$ and $k < h$, since each r_i is a subtree of some t_p such that $n - 1 < p$. But since $|r_1| < |t_n|$, this contradicts the fact that t is a minimal bad sequence. Hence, $\mathcal{D}$ is a *wqo*.

By Higman's theorem (theorem 3.2), the string embedding relation $\ll$ extending the preorder $\preceq$ on $\mathcal{D}$ is a *wqo* on $\mathcal{D}^*$. Hence, considering the infinite sequence over $\mathcal{D}^*$

$$\langle \langle s_{\varphi(1)}/1, s_{\varphi(1)}/2, \ldots, s_{\varphi(1)}/rank(s_{\varphi(1)}) \rangle, \ldots, \langle s_{\varphi(j)}/1, s_{\varphi(j)}/2, \ldots, s_{\varphi(j)}/rank(s_{\varphi(j)}) \rangle, \ldots \rangle,$$

there exist some $i, j > 0$ such that, letting $m = rank(s_{\varphi(i)})$ and $n = rank(s_{\varphi(j)})$,

$$\langle s_{\varphi(i)}/1, s_{\varphi(i)}/2, \ldots, s_{\varphi(i)}/m \rangle \ll \langle s_{\varphi(j)}/1, s_{\varphi(j)}/2, \ldots, s_{\varphi(j)}/n \rangle,$$

that is, there are some positive integers $j_1 < j_2 < \ldots < j_{m-1} < j_m \leq n$ such that

$$s_{\varphi(i)}/1 \preceq s_{\varphi(j)}/j_1, \ \ldots, \ s_{\varphi(i)}/m \preceq s_{\varphi(j)}/j_m.$$

Since we also have $f_{\varphi(i)} \sqsubseteq f_{\varphi(j)}$, by clause (3) of the definition of $\preceq$, we have $s_{\varphi(i)} \preceq s_{\varphi(j)}$. But s is a subsequence of t, and this contradicts the fact that t is bad. Hence, $\preceq$ is a *wqo* on T_Σ. $\square$

The above proof is basically due to Nash-Williams.

4.2 Kruskal's Theorem, Version 2

Another version of Kruskal's theorem that assumes a given preorder on T_Σ (and not just Σ) can also be proved. This version (found in J.J. Levy's unpublished notes [33]) can be used to show that certain orderings on trees are well-founded.

Definition 4.6 Assume that $\sqsubseteq$ is a preorder on T_Σ. The preorder $\preceq$ on T_Σ is defined inductively as follows: Either

(1) $f \preceq g(t_1, \ldots, t_n)$ iff $f \sqsubseteq g(t_1, \ldots, t_n)$; or

(2) $s \preceq g(\ldots, t, \ldots)$ iff $s \preceq t$; or

(3) $s = f(s_1, \ldots, s_m) \preceq g(t_1, \ldots, t_n) = t$ iff $s \sqsubseteq t$, and there exist some integers $j_1, \ldots, j_m$ such that $1 \leq j_1 < j_2 < \ldots < j_{m-1} < j_m \leq n$, $1 \leq m \leq n$, and

$$s_1 \preceq t_{j_1}, \ \ldots, \ s_m \preceq t_{j_m}.$$

It is easy to show that $\preceq$ is a preorder. It can also be shown that $\preceq$ is a partial order if $\sqsubseteq$ is a partial order. Again, (1) can be viewed as the special case of (3) for which $m = 0$ and, $n = 0$ is possible. It is also easy to see that $\preceq$ can be defined as the least preorder satisfying the following properties:

(1) $s \preceq f(\ldots, s, \ldots)$;

(2) $s = f(s_1, \ldots, s_m) \preceq g(t_1, \ldots, t_n) = t$ whenever $s \sqsubseteq t$ and there exist some integers $j_1, \ldots, j_m$ such that $1 \leq j_1 < j_2 < \ldots < j_{m-1} < j_m \leq n$, $1 \leq m \leq n$, and

$$s_1 \preceq t_{j_1}, \ \ldots, \ s_m \preceq t_{j_m}.$$

We can now prove another version of Kruskal's theorem.

Theorem 4.7 (J.J. Levy) If $\sqsubseteq$ is a *wqo* on T_Σ, then $\preceq$ is a *wqo* on T_Σ.

Proof. Assume that $\preceq$ is not a *wqo* on T_Σ. As in the proof of theorem 4.5, we find a minimal bad sequence t of elements of T_Σ.

Since $\sqsubseteq$ is a *wqo*, there is some infinite subsequence $t' = (t_{\psi(i)})_{i \geq 1}$ of t such that $t_{\psi(i)} \sqsubseteq t_{\psi(i+1)}$ for all $i \geq 1$. We claim that $|t_{\psi(i)}| \geq 2$ for all but finitely many $i \geq 1$. Otherwise, the sequence of one-node trees in t' must be infinite, and since $\sqsubseteq$ is a *wqo*, by clause (1) of the definition of $\preceq$, there are $i, j > 0$ such that $\psi(i) < \psi(j)$ and $t_{\psi(i)} \preceq t_{\psi(j)}$, contradicting the fact that t is bad.

Let $s = (t'_{\eta(i)})_{i \geq 1}$ be the infinite subsequence of t' consisting of all trees having at least two nodes. Since s is a subsequence of t' and t' is a subsequence of t, s is a subsequence of t of the form $s = (t_{\varphi(i)})_{i \geq 1}$ for some strictly monotonic function φ. Let

$$\mathcal{D} = \{t_{\varphi(i)}/j \mid i \geq 1, \ 1 \leq j \leq rank(t_{\varphi(i)})\}.$$

As in the proof of theorem 4.5, we can show that $\preceq$ is a *wqo* on $\mathcal{D}$.

By Higman's theorem (theorem 3.2), the string embedding relation $\ll$ extending the preorder $\preceq$ on $\mathcal{D}$ is a *wqo* on $\mathcal{D}^*$. Hence, considering the infinite sequence over $\mathcal{D}^*$

$$\langle\langle t_{\varphi(1)}/1, t_{\varphi(1)}/2, \ldots, t_{\varphi(1)}/rank(t_{\varphi(1)})\rangle, \ldots, \langle t_{\varphi(j)}/1, t_{\varphi(j)}/2, \ldots, t_{\varphi(j)}/rank(t_{\varphi(j)})\rangle, \ldots\rangle,$$

there exist some $i, j > 0$ such that, letting $m = rank(t_{\varphi(i)})$ and $n = rank(t_{\varphi(j)})$,

$$\langle t_{\varphi(i)}/1, t_{\varphi(i)}/2, \ldots, t_{\varphi(i)}/m\rangle \ll \langle t_{\varphi(j)}/1, t_{\varphi(j)}/2, \ldots, t_{\varphi(j)}/n\rangle,$$

that is, there are some positive integers $j_1 < j_2 < \ldots < j_{m-1} < j_m \leq n$ such that

$$t_{\varphi(i)}/1 \preceq t_{\varphi(j)}/j_1, \ \ldots, \ t_{\varphi(i)}/m \preceq t_{\varphi(j)}/j_m.$$

Since we also have $t_{\varphi(i)} \sqsubseteq t_{\varphi(j)}$ (because $s = (t_{\varphi(i)})_{i\geq 1}$ is also a subsequence of $t' = (t_{\psi(i)})_{i\geq 1}$ and $t_{\psi(i)} \sqsubseteq t_{\psi(i+1)}$ for all $i \geq 1$), by clause (3) of the definition of $\preceq$, we have $t_{\varphi(i)} \preceq t_{\varphi(j)}$. But this contradicts the fact that t is bad. Hence, $\preceq$ is a *wqo* on T_Σ. □

This second version of Kruskal's theorem (theorem 4.7) actually implies the first version (theorem 4.5). Indeed, if $\sqsubseteq$ is a preorder on Σ, we can extend it to a preorder on T_Σ by requiring that $s \sqsubseteq t$ iff $root(s) \sqsubseteq root(t)$. It is easy to check that with this definition of $\sqsubseteq$, definition 4.6 reduces to 4.4, and that theorem 4.7 is indeed theorem 4.5.

Kruskal's theorem has been generalized in a number of ways. Among these generalizations, we mention some versions using unavoidable sets of trees due to Puel [43, 44], and a version using well rewrite orderings due to Lescanne [30].

4.3 WQO's and Well-Founded Preorders

This second version of Kruskal's theorem also has the following applications. Recall that from lemma 2.4 a *wqo* is well-founded. The following proposition is very useful to prove that orderings on trees are well-founded.

Proposition 4.8 Let $\ll$ be a preorder on T_Σ and let $\leq$ be another preorder on T_Σ such that:

(1) If $f \ll g(t_1, \ldots, t_n)$, then $f \leq g(t_1, \ldots, t_n)$;

(2) $s \leq f(\ldots, s, \ldots)$;

(3) If $f(s_1, \ldots, s_m) \ll g(t_1, \ldots, t_n)$, and $s_1 \leq t_{j_1}, \ldots, s_m \leq t_{j_m}$ for some $j_1, \ldots, j_m$ such that $1 \leq j_1 < \ldots < j_m \leq n$, then $f(s_1, \ldots, s_m) \leq g(t_1, \ldots, t_n)$.

If $\ll$ is a *wqo*, then $\leq$ is a *wqo*.

Proof. Let $\preceq$ be the preorder associated with $\ll$ as in definition 4.6. Then, an easy induction shows that the conditions of the proposition imply that $\preceq \subseteq \leq$. By theorem 4.7, since $\ll$ is a *wqo*, $\preceq$ is also a *wqo*, which implies that $\leq$ is a *wqo*. By lemma 2.4, $\leq$ is well-founded. □

The following proposition also gives a sufficient condition for a preorder on trees to be well-founded.

Proposition 4.9 Assume Σ is finite, and let $\leq$ be a preorder on T_Σ satisfying the following conditions:

(1) $s \leq f(\ldots, s, \ldots)$;

(2) $s \leq t$ implies that $f(\ldots, s, \ldots) \leq f(\ldots, t, \ldots)$;

(3) $f(\ldots) \leq f(\ldots, s, \ldots)$.

Then, $\leq$ is well-founded.

Proof. Let $\ll$ be the preorder on T_Σ defined such that $s \ll t$ iff $root(s) = root(t)$. Since Σ is finite, $\ll$ is a *wqo*. Since it is clear that $\ll$ and $\leq$ satisfy the conditions of proposition 4.8, $\leq$ is well-founded. □

Proposition 4.8 can be used to show that certain orderings on trees are well-founded. These orderings play a crucial role in proving the termination of systems of rewrite rules and the correctness of Knuth-Bendix completion procedures. An introduction to the theory of these orderings will be presented in section 11, and for more details, the reader is referred to the comprehensive survey by Dershowitz [7] and to Dershowitz's fundamental paper [8].

It is natural to ask whether there is an analogue to Kruskal's theorem with respect to well-founded preorders instead of *wqo*. Indeed, it is possible to prove such a theorem, using Kruskal's theorem.

Theorem 4.10 If $\sqsubseteq$ is a well-founded preorder on T_Σ, then $\preceq$ is well-founded on T_Σ.

Proof. The proof is implicit in Levy [33], Dershowitz [8], and Lescanne [29]. Unfortunately, one cannot directly apply theorem 4.7, since $\sqsubseteq$ is not necessarily a *wqo*. However, there is a way around this problem. We use the fact that every well-founded preorder $\sqsubseteq$ can be extended to a total well-founded preorder $\leq$. This fact can be proved rather simply using Zorn's lemma. The point is that $\leq$ being total and well-founded is also a *wqo*. Now, we can apply theorem 4.7 since $\leq$ is a *wqo* on T_Σ, and so $\preceq_\leq$ is a *wqo* on T_Σ, and thus it is well-founded. Finally, we note that $\preceq_\leq$ contains $\preceq$, which proves that $\preceq$ is well-founded. □

Exercise: Find a proof of theorem 4.10 that does not use Zorn's lemma nor Kruskal's theorem.

4.4 Kruskal's Theorem, A Special Version

Kruskal's tree theorem is a very powerful theorem, and we state more interesting consequences. We consider the case where Σ is a finite set of symbols.

Definition 4.11 The preorder $\preceq$ on T_Σ is defined inductively as follows: Either

(1) $f \preceq f(t_1, \ldots, t_n)$, for every $f \in \Sigma$; or

(2) $s \preceq f(\ldots, t, \ldots)$ iff $s \preceq t$; or

(3) $f(s_1, \ldots, s_m) \preceq f(t_1, \ldots, t_n)$ iff $1 \leq m \leq n$, and there exist some integers $j_1, \ldots, j_m$ such that $1 \leq j_1 < j_2 < \ldots < j_{m-1} < j_m \leq n$ and

$$s_1 \preceq t_{j_1}, \ \ldots, \ s_m \preceq t_{j_m}.$$

Again, (1) can be viewed as the special case of (3) in which $m = 0$. For example,

$$f(f(h,h), h(a,b)) \preceq h(\mathbf{f}(g(\mathbf{f}(\mathbf{h}(b), a, \mathbf{h}(b))), g(a), h(\mathbf{h}(\mathbf{a}, \mathbf{b}, c)))).$$

It is also easy to show that the preorder $\preceq$ can be defined as the least preorder satisfying the following properties:

(1) $s \preceq f(\ldots, s, \ldots)$;

(2) $f(\ldots) \preceq f(\ldots, s, \ldots)$;

(3) $f(\ldots, s, \ldots) \preceq f(\ldots, t, \ldots)$ whenever $s \preceq t$.

Kruskal's theorem implies the following result.

Theorem 4.12 Given a finite alphabet Σ, $\preceq$ is a *wqo* on T_Σ.

Proof. Since any preorder on a finite set is a *wqo*, the identity relation on Σ is a *wqo*. But then, it is trivial to verify that the preorder $\preceq$ of definition 4.11 is obtained by specializing $\sqsubseteq$ to the identity relation in definition 4.4. Hence, the theorem is direct a consequence of theorem 4.5. □

In particular, when Σ consists of a single symbol, we have the well-known version of Kruskal's theorem on unlabeled trees [28], except that in Kruskal's paper, the notion of embedding is defined as a certain kind of function between tree domains. We find it more convenient to define the preorder $\preceq$ inductively, as in definition 4.4. For the sake of completeness, we present the alternate definition used by Simpson [47].

4.5 Tree Domains And Embeddings: An Alternate Definition

First, given a partial order $\leq$ on a set A, given any nonempty subset S of A, we say that $\leq$ is a *total order* on S iff for all $x, y \in S$, either $x \leq y$, or $y \leq x$. We also say that S is a *chain* (under $\leq$).

Definition 4.13 A finite tree domain is a nonempty set D together with a partial order $\leq$ satisfying the following properties:

(1) D has a least element $\perp$ (with respect to $\leq$).

(2) For every $x \in D$, the set $anc(x) = \{y \in D \mid y \leq x\}$ of ancestors of x is a chain under $\leq$.

Clearly $\perp$ corresponds to the root of the tree, and for every $x \in D$, the set $anc(x) = \{y \in D \mid y \leq x\}$ is the set of nodes in the unique path from the root to x. The main difference between definition 4.1 and definition 4.13 is that independent nodes of a tree domain as defined in definition 4.13 are *unordered*, and, in particular, the immediate successors of a node are unordered.

Given any two elements $x, y \in D$, the greatest element of the set $anc(x) \cap anc(y)$ is the greatest lower bound of x and y, and it is denoted as $x \wedge y$. It is the "lowest" common ancestor of x and y. A (labeled) tree is defined as in definition 4.2, but using definition 4.13 for that of a tree domain. The notion of an *embedding* (or *homeomorphic embedding*) is then defined as follows. Let Σ be a set with some preorder $\sqsubseteq$.

Definition 4.14 Given any two trees t_1 and t_2 with tree domains $\langle D_1, \leq_1 \rangle$ and $\langle D_2, \leq_2 \rangle$, an *embedding* h from t_1 to t_2 is an injective function $h : \langle D_1, \leq_1 \rangle \to \langle D_2, \leq_2 \rangle$ such that:

(1) $h(x \wedge y) = h(x) \wedge h(y)$, for all $x, y \in D_1$.

(2) $t_1(x) \sqsubseteq t_2(h(x))$, for every $x \in D_1$.

It is easily shown that h is monotonic (choose x, y such that $x \leq_1 y$). One can verify that when the immediate successors of a node are ordered, definition 4.4 is equivalent to definition 4.14.

Next, we shall consider an extremely interesting version of Kruskal's theorem due to Harvey Friedman. A complete presentation of this theorem and its ramifications is given by Simpson [47].

5 Friedman's Finite Miniaturization of Kruskal's Theorem

Friedman's version of Kruskal's theorem, which has been called a finite miniaturization of Kruskal's theorem, is remarkable from a proof-theoretic point of view because it is *not* provable in relatively strong logical systems. Actually, Kruskal's original theorem is also not provable in relatively strong logical systems, but Kruskal's version is a second-order statement (a Π^1_1 statement, meaning that it is of the form $\forall X A$, where X is a second-order variable ranging over infinite sequences and A is first-order), whereas Friedman's version is a first-order statement (a Π^0_2 statement, meaning that it is of the form $\forall x \exists y A$, where A only contains bounded first-order quantifiers).

From now on, we assume that Σ is a finite alphabet, and we consider the embedding preorder of definition 4.11.

Theorem 5.1 (Friedman) Let Σ be a finite set. For every integer $k \geq 1$, there exists some integer $n \geq 2$ so large that, for any finite sequence $\langle t_1, \ldots, t_n \rangle$ of trees in T_Σ with $|t_m| \leq k(m+1)$ for all m, $1 \leq m \leq n$, there exist some integers i, j such that $1 \leq i < j \leq n$ and $t_i \preceq t_j$.

Proof. Following the hint given by Simpson [47], we give a proof using theorem 4.12 and König's lemma. Assume that the theorem fails. Let us say that a finite sequence $\langle t_1, \ldots, t_n \rangle$ such that $|t_m| \leq k(m+1)$ for all m, $1 \leq m \leq n$, is *good* iff there exist some integers i, j such that $1 \leq i < j \leq n$ and $t_i \preceq t_j$, and otherwise, that it is *bad*. Then, there is some $k \geq 1$ such that for all $n > 1$, there is some bad sequence $\langle t_1, \ldots, t_n \rangle$ (and $|t_m| \leq k(m+1)$ for all m, $1 \leq m \leq n$). Observe that any initial subsequence $\langle t_1, \ldots, t_j \rangle$, $j < n$, of a bad sequence is also bad. Furthermore, the size restriction ($|t_m| \leq k(m+1)$ for all m, $1 \leq m \leq n$) and the fact that Σ is finite implies that there are only finitely many bad sequences of length n. Hence, the set of finite bad sequences can be arranged into an infinite tree $\mathcal{T}$ as follows: the root of $\mathcal{T}$ is the empty sequence, and every finite bad sequence t is connected to the root by the unique path consisting of all the initial subsequences of t. From our previous remark, this infinite tree is finite-branching. By König's lemma, this tree contains an infinite path s. But since all finite initial subsequences of s are bad, s itself is bad, and this contradicts theorem 4.12. □

A stronger version of the previous theorem also due to Friedman holds.

Theorem 5.2 (Friedman) Let Σ be a finite set. For every integer $k \geq 2$, there exists some integer $n \geq 2$ so large that, for any finite sequence $\langle t_1, \ldots, t_n \rangle$ of trees in T_Σ with $|t_m| \leq m$ for all m, $1 \leq m \leq n$, there exist some integers $i_1, \ldots, i_k$ such that $1 \leq i_1 < \ldots < i_k \leq n$ and $t_{i_1} \preceq \ldots \preceq t_{i_k}$.

Proof. The proof is very similar to that of theorem 5.1, but lemma 2.5 also needs to be used at the end. □

Note that theorems 5.1 and 5.2 are both of the form $\forall k \exists n A(k,n)$, where $A(k,n)$ only contains bounded quantifiers, that is, they are Π^0_2 statements. Hence, each statement defines a function Fr, where $Fr(k)$ is the least integer n such that $\forall k \exists n A(k,n)$ holds.

One may ask how quickly this function grows. Is it exponential, super exponential, or worse? Well, this function grows extremely fast. It grows faster than Ackermann's function, and, even though it is recursive, it is not *provably total recursive* in fairly strong logical theories, including Peano's arithmetic. We will consider briefly hierarchies of fast-growing functions in section 12. For more details, we refer the reader to Cichon and Wainer [4], Wainer [54], and to Smoryński's articles [50,51].

The other remarkable property of the two previous theorems is that neither is provable in fairly strong logical theories (ATR_0, see section 10). The technical reason is that it is possible to define a function mapping finite trees to (rather large) countable ordinals, and this function is order preserving (between the embedding relation $\preceq$ on trees and the ordering relation on ordinals). This is true in particular for the ordinal Γ_0 (see Schütte [46], chapters 13, 14). For further details, see the articles by Simpson and Smoryński in [21]. We shall present the connection with Γ_0 in sections 9 and 10.

6 The Countable Ordinals

In this section, we gather some definitions and results about the countable ordinals needed to explain what Γ_0 is. This ordinal plays a central role in proof theoretic investigations of a subsystem of second-order arithmetic known as "predicative analysis", which has been studied extensively by Feferman [13] and Schütte [46]. Schütte's axiomatic presentation of the countable ordinals ([46], chapters 13, 14) is particularly convenient (and elegant), and we follow it. Most proofs are omitted. They can be found in Schütte [46].

6.1 A Preview of Γ_0

Proof theorists use (large) ordinals in inductive proofs establishing the consistency of certain theories. In order for these proofs to be as constructive as possible, it is crucial to describe these ordinals using systems of constructive ordinal notations. One way to obtain constructive ordinal notation systems is to build up inductively larger ordinals from smaller ones using functions on the ordinals. For example, if $\mathcal{O}$ denotes the set of countable ordinals, it is possible to define two functions $+$ and $\alpha \mapsto \omega^\alpha$ (where ω is the least infinite ordinal) generalizing addition and exponentiation on the natural numbers. Due to a result of Cantor, for every ordinal $\alpha \in \mathcal{O}$, if $\alpha > 0$, there are unique ordinals $\alpha_1 \geq \ldots \geq \alpha_n$, $n \geq 1$, such that

$$\alpha = \omega^{\alpha_1} + \cdots + \omega^{\alpha_n}. \qquad (*)$$

This suggests a constructive ordinal notation system. Define $\mathcal{C}$ to be the smallest set of ordinals containing 0 and closed under $+$ and $\alpha \mapsto \omega^\alpha$.

Do we have $\mathcal{C} = \mathcal{O}$? The answer is *no*. Indeed, strange things happen with infinite ordinals. For some ordinals α, β such that $0 < \alpha < \beta$, we can have $\alpha + \beta = \beta$, and even $\omega^\alpha = \alpha$!

An ordinal $\beta > 0$ such that $\alpha + \beta = \beta$ for all $\alpha < \beta$ is called an *additive principal ordinal*. It can be shown that an ordinal is an additive principal ordinal iff it is of the form ω^η for some η.

The general phenomenon that we are witnessing is the fact that if a function $f : \mathcal{O} \to \mathcal{O}$ satisfies a certain continuity condition, then it has *fixed points* (an ordinal α is a fixed point of f iff $f(\alpha) = \alpha$).

The least ordinal such that $\omega^\alpha = \alpha$ (the least fixed point of $\alpha \mapsto \omega^\alpha$) is denoted by ϵ_0, and $\mathcal{C}$ provides a constructive ordinal notation system for the ordinals $< \epsilon_0$. The main point here, is that for every ordinal $\alpha < \epsilon_0$, we can guarantee that $\alpha_i < \alpha$ in the decomposition $(*)$.

Unfortunately ϵ_0 is too small for our purpose (which is to relate the embedding relation $\preceq$ on finite trees with the ordering on Γ_0). To go beyond ϵ_0, we need functions more powerful than $\alpha \mapsto \omega^\alpha$. Such a hierarchy $(\varphi_\alpha)_{\alpha \in \mathcal{O}}$ can be defined inductively, starting from $\alpha \mapsto \omega^\alpha$.

We let φ_0 be the function $\alpha \mapsto \omega^\alpha$, and for every $\alpha > 0$, $\varphi_\alpha : \mathcal{O} \to \mathcal{O}$ enumerates the common fixed points of the functions φ_β, for all $\beta < \alpha$ (the ordinals η such that $\varphi_\beta(\eta) = \eta$ for all $\beta < \alpha$).

Then, we have a function $\varphi : \mathcal{O} \times \mathcal{O} \to \mathcal{O}$, defined such that $\varphi(\alpha, \beta) = \varphi_\alpha(\beta)$ for all $\alpha, \beta \in \mathcal{O}$. Note, $\varphi(1, 0) = \epsilon_0$!

The function φ has lots of fixed points. We can have $\varphi(\alpha, \beta) = \beta$, in which case β is called an *α-critical ordinal*, or $\varphi(\alpha, 0) = \alpha$ (but we can't have $\varphi(\alpha, \beta) = \alpha$ for $\beta > 0$). Ordinals such that $\varphi(\alpha, 0) = \alpha$ are called *strongly critical*.

It can be shown that for every additive principal ordinal $\gamma = \omega^\eta$, there exist unique α, β with $\alpha \leq \gamma$ and $\beta < \gamma$, such that $\gamma = \varphi(\alpha, \beta)$. But we can't guarantee that $\alpha < \gamma$, because $\varphi(\alpha, 0) = \alpha$ when α is a strongly critical ordinal. This is where Γ_0 comes in!

The ordinal Γ_0 is the *least* ordinal such that $\varphi(\alpha, 0) = \alpha$ (the least strongly critical ordinal). It can be shown that for all $\alpha, \beta < \Gamma_0$, we have $\alpha + \beta < \Gamma_0$ and $\varphi(\alpha, \beta) < \Gamma_0$, and also that for every additive principal ordinal $\gamma < \Gamma_0$, $\gamma = \varphi(\alpha, \beta)$ for unique ordinals such that *both* $\alpha < \gamma$ and $\beta < \gamma$. This fact together with the Cantor normal form $(*)$ yields a constructive ordinal notation system for the ordinals $< \Gamma_0$ described in the sequel.

The reason why we were able to build the hierarchy $(\varphi_\alpha)_{\alpha \in \mathcal{O}}$ is that these functions satisfy certain conditions: they are increasing and continuous. Such functions are called *normal functions*. What is remarkable is that the function $\varphi(-, 0)$ is also a normal function, and so, it is possible to repeat the previous hierarchy construction, but this time, starting from $\varphi(-, 0)$. But there is no reason to stop there, and we can continue on and on . . .!

We have what is called a *Veblen hierarchy* [53]. However, this is going way beyond the scope of these notes (transfinitely beyond!). The intrigued reader is referred to a paper by Larry Miller [34].

6.2 Axioms for the Countable Ordinals

Recall that a set A is countable iff either $A = \emptyset$ or there is a surjective (onto) function $f : \mathbf{N} \to A$ with domain $\mathbf{N}$, the set of natural numbers. In particular, every finite set is countable.

Given a set A and a partial order $\leq$ on A, we say that A is *well-ordered* by $\leq$ iff every nonempty subset of A has a least element.

This definition implies that a well-ordered set is totally ordered. Indeed, every subset $\{x, y\}$ of A consisting of two elements has a least element, and so, either $x \leq y$ or $y \leq x$.

We say that a subset $S \subseteq A$ of A is *strictly bounded* iff there is some $b \in A$ such that $x < b$ for all $x \in S$ (recall that $x < y$ iff $x \leq y$ and $x \neq y$). A subset S of A that is not strictly bounded is called *unbounded*. The set of countable ordinals is defined by the following axioms.

Definition 6.1 A set $\mathcal{O}$ together with a partial order $\leq$ on $\mathcal{O}$ satisfies the axioms for the countable ordinals iff the following properties hold:

(1) $\mathcal{O}$ is well-ordered by $\leq$.

(2) Every strictly bounded subset of $\mathcal{O}$ is countable.

(3) Every countable subset of $\mathcal{O}$ is strictly bounded.

Applying axiom (3) to the empty set (which is a subset of $\mathcal{O}$), we see that $\mathcal{O}$ is nonempty. Applying axiom (1) to $\mathcal{O}$, we see that $\mathcal{O}$ has a least element denoted by 0. Repeating this argument, we see that $\mathcal{O}$ is infinite. However, $\mathcal{O}$ is *not* countable. Indeed if $\mathcal{O}$ was countable, by axiom (3), there would be some $\alpha \in \mathcal{O}$ such that $\beta < \alpha$ for all $\beta \in \mathcal{O}$, which implies $\alpha < \alpha$, a contradiction.

It is possible to show that axioms (1)-(3) define the set of countable ordinals up to isomorphism. From now on, the elements of the set $\mathcal{O}$ will be called ordinals (strictly speaking, they should be called countable ordinals).

Given a property $P(x)$ of the set of countable ordinals, the principle of *transfinite induction* is the following:

- If $P(0)$ holds, and
- for every $\alpha \in \mathcal{O}$ such that $\alpha > 0$, $\forall\beta(\beta < \alpha \supset P(\beta))$ implies $P(\alpha)$, then
- $P(\gamma)$ holds for all $\gamma \in \mathcal{O}$.

We have the following fundamental metatheorem.

Theorem 6.2 The principle of transfinite induction is valid for $\mathcal{O}$.

Proof. Assume that the principle of transfinite induction does not hold. Then, $P(0)$ holds, for every $\alpha \in \mathcal{O}$ such that $\alpha > 0$, $\forall\beta(\beta < \alpha \supset P(\beta))$ implies $P(\alpha)$, but the set $W = \{\alpha \in \mathcal{O} \mid P(\alpha) = \textbf{false}\}$ is nonempty. By axiom (1), this set has a least element γ. Clearly, $\gamma \neq 0$, and $P(\beta)$ must hold for all $\beta < \gamma$, since otherwise γ would not be the least

element of W. Hence, $\forall \beta < \gamma P(\beta)$ holds, and from above, this implies that $P(\gamma)$ holds, contradicting the definition of γ. $\square$

By axioms (1) and (3), for every ordinal α, there is a smallest ordinal β such that $\alpha < \beta$. Indeed, the set $\{\alpha\}$ is countable, hence by axiom (3) the set $\{\beta \in \mathcal{O} \mid \alpha < \beta\}$ is nonempty, and by axiom (1), it has a least element. This ordinal is denoted by α', and is called the *successor* of α. We have the following properties:

$$\alpha < \alpha'$$
$$\alpha < \beta \Rightarrow \alpha' \leq \beta$$
$$\alpha < \beta' \Rightarrow \alpha \leq \beta.$$

An ordinal β is called a *successor ordinal* iff there is some $\alpha \in \mathcal{O}$ such that $\beta = \alpha'$. A *limit ordinal* is an ordinal that is neither 0 nor a successor ordinal.

Given any countable subset $M \subseteq \mathcal{O}$, by axiom (3), the set $\{\alpha \in \mathcal{O} \mid \forall \beta \in M(\beta \leq \alpha)\}$ is nonempty, and by axiom (1), it has a least element. This ordinal denoted by $\bigsqcup M$ is the *least upper bound* of M, and it satisfies the following properties:

$$\alpha \in M \Rightarrow \alpha \leq \bigsqcup M$$
$$\alpha \leq \beta \text{ for all } \alpha \in M \Rightarrow \bigsqcup M \leq \beta$$
$$\beta < \bigsqcup M \Rightarrow \exists \alpha \in M \text{ such that } \beta < \alpha.$$

We have the following propositions.

Proposition 6.3 If M is a nonempty countable subset of $\mathcal{O}$ and M has no maximal element, then $\bigsqcup M$ is a limit ordinal.

Proposition 6.4 For all $\alpha, \beta \in \mathcal{O}$, if $\gamma < \beta$ for all $\gamma < \alpha$, then $\alpha \leq \beta$.

Proof. The proposition is clear if $\alpha = 0$. If α is a successor ordinal, $\alpha = \delta'$ for some δ, and since $\delta < \alpha$, by the hypothesis we have $\delta < \beta$, which implies $\alpha = \delta' \leq \beta$. If α is a limit ordinal, we prove that $\alpha = \bigsqcup\{\gamma \in \mathcal{O} \mid \gamma < \alpha\}$, which implies that $\alpha \leq \beta$, since by the hypothesis β is an upper bound of the set $\{\gamma \in \mathcal{O} \mid \gamma < \alpha\}$. Let $\delta = \bigsqcup\{\gamma \in \mathcal{O} \mid \gamma < \alpha\}$. First, it is clear that α is an upper bound of the set $\{\gamma \in \mathcal{O} \mid \gamma < \alpha\}$, and so $\delta \leq \alpha$. If $\delta < \alpha$, since α is a limit ordinal, we have $\delta' < \alpha$, contradicting the fact that δ is the least upper bound of the set $\{\gamma \in \mathcal{O} \mid \gamma < \alpha\}$. Hence, $\delta = \alpha$. $\square$

Definition 6.5 The set $\mathbf{N}$ of *finite ordinals* is the smallest subset of $\mathcal{O}$ that contains 0 and is closed under the successor function.

It is not difficult to show that $\mathbf{N}$ is countable and has no maximal element. The least upper bound of $\mathbf{N}$ is denoted by ω.

Proposition 6.6 The ordinal ω is the least limit ordinal. For every $\alpha \in \mathcal{O}$, $\alpha < \omega$ iff $\alpha \in \mathbf{N}$.

It is easy to see that limit ordinals satisfy the following property: For every limit ordinal β

$$\alpha < \beta \Rightarrow \alpha' < \beta.$$

6.3 Ordering Functions

Given any ordinal $\alpha \in \mathcal{O}$, let $\mathcal{O}(\alpha)$ be the set $\{\beta \in \mathcal{O} \mid \beta < \alpha\}$. Clearly, $\mathcal{O}(0) = \emptyset$, $\mathcal{O}(\omega) = \mathbf{N}$, and by axiom (2), each $\mathcal{O}(\alpha)$ is countable.

Definition 6.7 A subset $A \subseteq \mathcal{O}$ is an *$\mathcal{O}$-segment* iff for all $\alpha, \beta \in \mathcal{O}$, if $\beta \in A$ and $\alpha < \beta$, then $\alpha \in A$.

The set $\mathcal{O}$ itself is an $\mathcal{O}$-segment, and an $\mathcal{O}$-segment which is a proper subset of $\mathcal{O}$ is called a *proper $\mathcal{O}$-segment*. It is easy to show that A is a proper $\mathcal{O}$-segment iff $A = \mathcal{O}(\alpha)$ for some $\alpha \in \mathcal{O}$.

We now come to the crucial concept of an ordering function.

Definition 6.8 Given a subset $B \subseteq \mathcal{O}$, a function $f : A \to B$ is an *ordering function* for B iff:

(1) The domain of f is an $\mathcal{O}$-segment.

(2) The function f is *strictly monotonic (or increasing)*, that is, for all $\alpha, \beta \in \mathcal{O}$, if $\alpha < \beta$, then $f(\alpha) < f(\beta)$.

(3) The range of f is B.

Intuitively speaking, an ordering function f of a set B *enumerates* the elements of the set B in increasing order. Observe that an ordering function f is bijective, since by (3), $f(A) = B$, and by (2), f is injective. Note that the ordering function for the empty set is the empty function. The following fundamental propositions are shown by transfinite induction.

Proposition 6.9 If $f : A \to B$ is an ordering function, then $\alpha \leq f(\alpha)$ for all $\alpha \in A$

Proof. Clearly, $0 \leq f(0)$. Given any ordinal $\alpha > 0$, for every $\beta < \alpha$, by the induction hypothesis, $\beta \leq f(\beta)$. Since f is strictly monotonic, $f(\beta) < f(\alpha)$. Hence, $\beta < f(\alpha)$ for all $\beta < \alpha$, and by proposition 6.4, this implies that $\alpha \leq f(\alpha)$. $\square$

Proposition 6.10 Every subset $B \subseteq \mathcal{O}$ has at most one ordering function $f : A \to B$.

Proof. Let $f_i : A_i \to B$, $i = 1, 2$, be two ordering functions for B. We show by transfinite induction that, if $\alpha \in A_1$, then $\alpha \in A_2$ and $f_1(\alpha) = f_2(\alpha)$. If $B = \emptyset$, then clearly $f_1 = f_2 : \emptyset \to \emptyset$. Otherwise, since A_1 and A_2 are $\mathcal{O}$-segments, $0 \in A_1$ and $0 \in A_2$. Since f_2 is surjective, there is some $\alpha \in A_2$ such that $f_2(\alpha) = f_1(0)$. By (strict) monotonicity of f_2, we have $f_2(0) \leq f_1(0)$. Similarly, since f_1 is surjective, there is some $\beta \in A_1$ such that $f_1(\beta) = f_2(0)$, and by (strict) monotonicity of f_1, we have $f_1(0) \leq f_2(0)$. Hence $f_1(0) = f_2(0)$. Now, assume $\alpha > 0$. Since f_2 is surjective, there is some $\beta \in A_2$ such that $f_2(\beta) = f_1(\alpha)$. If $\beta < \alpha$, since A_1 is an $\mathcal{O}$-segment, $\beta \in A_1$, and by the induction hypothesis, $\beta \in A_2$ and $f_1(\beta) = f_2(\beta)$. By strict monotonicity, $f_2(\beta) = f_1(\beta) < f_1(\alpha)$, a contradiction.

Hence, $\beta \geq \alpha$, and since A_2 is an $\mathcal{O}$-segment and $\beta \in A_2$, we have $\alpha \in A_2$. Assume $\beta > \alpha$. By strict monotonicity, $f_2(\alpha) < f_2(\beta)$. Since f_1 is surjective, there is some $\gamma \in A_1$ such that $f_1(\gamma) = f_2(\alpha)$. Since $f_2(\alpha) = f_1(\gamma)$, $f_2(\beta) = f_1(\alpha)$, and $f_2(\alpha) < f_2(\beta)$, we have $f_1(\gamma) < f_1(\alpha)$. By strict monotonicity, we have $\gamma < \alpha$. By the induction hypothesis, $f_1(\gamma) = f_2(\gamma)$, and since $f_1(\gamma) = f_2(\alpha)$, then $f_2(\gamma) = f_2(\alpha)$. Since f_2 is injective, we have $\alpha = \gamma$, a contradiction. Hence, $\alpha = \beta$ and $f_1(\alpha) = f_2(\alpha)$. Therefore, we have shown that $A_1 \subseteq A_2$ and for every $\alpha \in A_1$, $f_1(\alpha) = f_2(\alpha)$. Using a symmetric argument, we can show that $A_2 \subseteq A_1$ and for every $\alpha \in A_2$, $f_1(\alpha) = f_2(\alpha)$. Hence, $A_1 = A_2$ and $f_1 = f_2$. □

Given a set $B \subseteq \mathcal{O}$, for every $\beta \in B$, let $B(\beta) = \{\gamma \in B \mid \gamma < \beta\}$. Sets of the form $B(\beta)$ are called *proper segments* of B. Observe that $B(\beta) = B \cap \mathcal{O}(\beta)$. Using proposition 6.10, we prove the following crucial result.

Proposition 6.11 Every subset $B \subseteq \mathcal{O}$ has a unique ordering function $f : A \to B$.

Proof. First, the following claim is shown.

Claim: If every proper segment $B(\beta)$ of a set $B \subseteq \mathcal{O}$ has an ordering function, then B has an ordering function.

Proof of claim. The idea is to construct a function $g : B \to \mathcal{O}$ and to show that g is strictly monotonic and that its range is an $\mathcal{O}$-segment. Then, the inverse of g is an ordering function for B. By the hypothesis, for every $\beta \in B$, we have an ordering function $f_\beta : A_\beta \to B(\beta)$ for each proper segment $B(\beta)$ of B. By axiom (2) (in definition 6.1), $B(\beta)$ is countable. Since f_β is bijective, A_β is also countable, and therefore, it is a proper $\mathcal{O}$-segment. Hence, for every $\beta \in B$, there is a unique ordinal γ such that $A_\beta = \mathcal{O}(\gamma)$, and we define the function $g : B \to \mathcal{O}$ such that $g(\beta) = \gamma$.

We show that g is strictly monotonic. Let $\beta_1 < \beta_2$, $\beta_1, \beta_2 \in B$. Since the function $f_{\beta_2} : \mathcal{O}(g(\beta_2)) \to B(\beta_2)$ is surjective and $\beta_1 \in B(\beta_2)$ (since $\beta_1 < \beta_2$ and $\beta_2 \in B$), there is

some $\alpha < g(\beta_2)$ such that $f_{\beta_2}(\alpha) = \beta_1$. Observe that the restriction of f_{β_2} to $\mathcal{O}(\alpha)$ is an ordering function of $B(\beta_1)$. Since $f_{\beta_1} : A_{\beta_1} \to B(\beta_1)$ is also an ordering function for $B(\beta_1)$, by proposition 6.10, $\mathcal{O}(\alpha) = \mathcal{O}(g(\beta_1))$, and therefore, $g(\beta_1) = \alpha < g(\beta_2)$.

We show that $g(B)$ is an $\mathcal{O}$-segment. We have to show that for every $\gamma \in g(B)$, if $\alpha < \gamma$, then $\alpha \in g(B)$. Let $\beta \in B$ such that $\gamma = g(\beta)$. Since $f_\beta : \mathcal{O}(g(\beta)) \to B(\beta)$ and $\alpha < g(\beta)$, $f_\beta(\alpha) = \beta_0$ for some $\beta_0 \in B(\beta)$. The restriction of f_β to $\mathcal{O}(\alpha)$ is an ordering function of $B(\beta_0)$. Since $f_{\beta_0} : \mathcal{O}(g(\beta_0)) \to B(\beta_0)$ is also an ordering function for $B(\beta_0)$, by proposition 6.10, $\alpha = g(\beta_0)$, and therefore $\alpha \in g(B)$.

Since the function $g : B \to \mathcal{O}$ is strictly monotonic and $g(B)$ is an $\mathcal{O}$-segment, say A, its inverse $g^{-1} : A \to B$ is an ordering function for B. This proves the claim. □

Let $B \subseteq \mathcal{O}$. For every $\beta \in B$, note that every proper segment of $B(\beta)$ is of the form $B(\beta_0)$ for some $\beta_0 < \beta$. Using the previous claim, it follows by transfinite induction that every proper segment $B(\beta)$ of B has an ordering function. By the claim, B itself has an ordering function. By proposition 6.10, this function is unique. □

An important property of ordering functions is continuity.

Definition 6.12 A subset $B \subseteq \mathcal{O}$ is *closed* iff for every countable nonempty set M,

$$M \subseteq B \Rightarrow \bigsqcup M \in B.$$

An ordering function $f : A \to B$ is *continuous* iff A is closed and for every nonempty countable set $M \subseteq A$,

$$f(\bigsqcup M) = \bigsqcup f(M).$$

Proposition 6.13 The ordering function $f : A \to B$ of a set B is continuous iff B is closed.

Proof. Let $f : A \to B$ be the ordering function of B. First, assume that f is continuous. Since f is bijective, for every nonempty countable subset $M \subseteq B$, there is some nonempty countable subset $U \subseteq A$ such that $f(U) = M$. Since f is continuous, $f(\bigsqcup U) = \bigsqcup f(U) = \bigsqcup M$, and therefore $\bigsqcup M \in f(A) = B$, and B is closed.

Conversely, assume that B is closed. Let $U \subseteq A$ be a nonempty countable subset of A. Since f is bijective, $f(U)$ is a nonempty countable subset of B. Since B is closed, $\bigsqcup f(U) \in B$. Since $B = f(A)$, there is some $\alpha \in A$ such that $f(\alpha) = \bigsqcup f(U)$. Since $f(\alpha) = \bigsqcup f(U)$, for every $\delta \in U$, we have $f(\delta) \leq f(\alpha)$, and by strict monotonicity of f, this implies that $\delta \leq \alpha$. Hence $\bigsqcup U \leq \alpha$. Since A is an $\mathcal{O}$-segment, $\bigsqcup U \in A$. Hence, A is closed. For all $\delta \in U$, $\delta \leq \bigsqcup U$, and so $f(\delta) \leq f(\bigsqcup U)$. Then, $f(\bigsqcup U)$ is an upper bound for

$f(U)$, and so $\bigsqcup f(U) \leq f(\bigsqcup U)$. Also, since $\bigsqcup U \leq \alpha$, we have $f(\bigsqcup U) \leq f(\alpha) = \bigsqcup f(U)$. But then, $\bigsqcup f(U) = f(\bigsqcup U)$, and f is continuous. □

An ordering function that is continuous and whose domain is the entire set $\mathcal{O}$ is called a *normal function*. Normal functions play a crucial role in the definition of Γ_0.

Proposition 6.14 The ordering function $f : A \to B$ of a set B is a normal function iff B is closed and unbounded.

Proof. By axiom (2) and (3) (in definition 6.1), a subset M of $\mathcal{O}$ is bounded iff it is countable. Since an ordering function $f : A \to B$ is bijective, it follows that B is unbounded iff A is unbounded. But A is an $\mathcal{O}$-segment, and $\mathcal{O}$ is the only unbounded $\mathcal{O}$-segment (since a proper $\mathcal{O}$-segment is bounded). Hence, the ordering function f has domain $\mathcal{O}$ iff B is unbounded. This together with proposition 6.13 yields proposition 6.14. □

We now show that normal functions have fixed points.

Proposition 6.15 Let $f : \mathcal{O} \to \mathcal{O}$ be a continuous function. For every $\alpha \in \mathcal{O}$, let $f^0(\alpha) = \alpha$, and $f^{n+1}(\alpha) = f(f^n(\alpha))$ for all $n \geq 0$. If $\alpha \leq f(\alpha)$ for every $\alpha \in \mathcal{O}$, then $\bigsqcup_{n\geq 0} f^n(\alpha)$ is the least fixed point of f that is $\geq \alpha$, and $\bigsqcup_{n\geq 0} f^n(\alpha')$ is the least fixed point of f that is $> \alpha$.

Proof. First, observe that a continuous function is monotonic, by applying the continuity condition to each set $\{\alpha, \beta\}$ with $\alpha \leq \beta$. Since f is continuous,

$$\begin{aligned} f(\bigsqcup_{n\geq 0} f^n(\alpha)) &= \bigsqcup_{n\geq 0} f(f^n(\alpha)) \\ &= \bigsqcup_{n\geq 0} f^{n+1}(\alpha) \\ &= \bigsqcup_{n\geq 1} f^n(\alpha) \\ &= \bigsqcup_{n\geq 0} f^n(\alpha), \end{aligned}$$

since $\alpha \leq f(\alpha)$. Hence, $\bigsqcup_{n\geq 0} f^n(\alpha)$ is a fixed point of f that is $\geq \alpha$. Let β be any fixed point of f such that $\alpha \leq \beta$. We show by induction that $f^n(\alpha) \leq \beta$. For $n = 0$, this follows from the fact that $f^0(\alpha) = \alpha$ and the hypothesis $\alpha \leq \beta$. If $f^n(\alpha) \leq \beta$, since f is monotonic we have, $f(f^n(\alpha)) \leq f(\beta)$, that is, $f^{n+1}(\alpha) \leq \beta$, since $f^{n+1}(\alpha) = f(f^n(\alpha))$ and $f(\beta) = \beta$ (because β is a fixed point of f). Hence, $\bigsqcup_{n\geq 0} f^n(\alpha) \leq \beta$, which shows that $\bigsqcup_{n\geq 0} f^n(\alpha)$ is the least fixed point of f that is $\geq \alpha$.

From above, $\bigsqcup_{n\geq 0} f^n(\alpha')$ is the least fixed point of f that is $\geq \alpha'$, and since $\beta \geq \alpha'$ iff $\beta > \alpha$, the second part of the lemma holds. □

Corollary 6.16 For every normal function f, for every $\alpha \in \mathcal{O}$, $\bigsqcup_{n\geq 0} f^n(\alpha)$ is the least fixed point of f that is $\geq \alpha$, and $\bigsqcup_{n\geq 0} f^n(\alpha')$ is the least fixed point of f that is $> \alpha$.

Proof. Since a normal function is continuous and $\alpha \leq f(\alpha)$ for all α, the corollary follows from proposition 6.15. □

Using the concept of a normal function, we are going to define addition and exponentiation of ordinals.

6.4 Addition and Exponentiation of Ordinals

For every $\alpha \in \mathcal{O}$, let $B_\alpha = \{\beta \in \mathcal{O} \mid \alpha \leq \beta\}$. Let f_α be the ordering function of B_α given by proposition 6.11. It is easy to see that B_α is closed and unbounded. Hence, by proposition 6.14, f_α is a normal function. We shall write $\alpha + \beta$ for $f_\alpha(\beta)$. The following properties of $+$ can be shown:

$\alpha \leq \alpha + \beta$.

$\beta < \gamma \Rightarrow \alpha + \beta < \alpha + \gamma$ (right strict monotonicity).

If $\alpha \leq \beta$, then there is a unique γ such that $\alpha + \gamma = \beta$.

For every limit ordinal $\beta \in \mathcal{O}$, $\bigsqcup \mathcal{O}(\beta) = \beta$, and $\alpha + \beta = \bigsqcup\{\alpha + \gamma \mid \gamma \in \mathcal{O}(\beta)\}$.

$\alpha + 0 = \alpha$.

$\alpha + \beta' = (\alpha + \beta)'$.

$\beta \leq \alpha + \beta$.

$0 + \beta = \beta$

$(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.

$\alpha \leq \beta \Rightarrow \alpha + \gamma \leq \beta + \gamma$ (left weak monotonicity).

It should be noted that addition of ordinals is *not* commutative. Indeed, $0' + \omega = \bigsqcup \mathbf{N} = \omega$, but $\omega < \omega + 0'$ by right strict monotonicity. Also,

Definition 6.17 An ordinal $\alpha \in \mathcal{O}$ is a *principal additive ordinal* iff $\alpha \neq 0$ and for every $\beta < \alpha$, $\beta + \alpha = \alpha$.

Clearly, $1 = 0'$ is the smallest additive principal ordinal, and it is not difficult to show that ω is the least additive principal ordinal greater than 1. Note that $\alpha + 1 = \alpha'$.

If α is an additive principal ordinal, then $\mathcal{O}(\alpha)$ is closed under addition.

Proposition 6.18 The set of additive principal ordinals is closed and unbounded.

Proof. First, we show unboundedness. Given any ordinal α, let $\beta_0 = \alpha'$, $\beta_{n+1} = \beta_n + \beta_n$, $M = \{\beta_n \mid n \in \mathbf{N}\}$, and $\beta = \bigsqcup M$. Since $\beta_0 = \alpha' > 0$, we have $\beta_n > 0$ for all $n \geq 0$, and by right strict monotonicity of $+$, $\beta_n < \beta_n + \beta_n = \beta_{n+1}$. Hence, $\alpha < \beta_n < \beta$ for all $n \geq 0$, and $\beta > 0$. If $\eta < \beta$, then there is some $n \geq 0$ such that $\eta < \beta_n$. Hence, for all $m \geq n$, $\eta + \beta_m \leq \beta_m + \beta_m = \beta_{m+1} < \beta$. Hence, $\bigsqcup\{\eta + \beta_n \mid n \in \mathbf{N}\} \leq \beta$. But we also have $\beta \leq \eta + \beta = \bigsqcup\{\eta + \beta_n \mid n \in \mathbf{N}\} \leq \beta$. Hence, $\eta + \beta = \beta$ for all $\eta < \beta$. Therefore, β is an additive principal ordinal.

Next, we show closure. Let M be a nonempty set of additive principal ordinals. Since for every $\beta \in M$, $\beta > 0$, we have $\bigsqcup M > 0$. Let $\eta < \bigsqcup M$. Then, there is some $\alpha \in M$ such that $\eta < \alpha$. For every $\beta \in M$, if $\beta \geq \alpha$, then $\eta < \beta$, and since β is additive principal, $\eta + \beta = \beta$. Hence, $\bigsqcup\{\eta + \beta \mid \beta \in M\} = \bigsqcup M$ for all $\eta < \bigsqcup M$, which shows that $\bigsqcup M$ is additive principal. □

By proposition 6.14, the ordering function of the set of additive principal ordinals is a normal function.

Definition 6.19 The ordering function of the set of additive principal ordinals is a normal function whose value for every ordinal α is denoted by ω^α.

The following properties hold.

$0 < \omega^\alpha$.

$\beta < \omega^\alpha \Rightarrow \beta + \omega^\alpha = \omega^\alpha$.

$\alpha < \beta \Rightarrow \omega^\alpha < \omega^\beta$.

For every additive principal ordinal β, there is some α such that $\beta = \omega^\alpha$.

For every limit ordinal β, $\omega^\beta = \bigsqcup\{\omega^\alpha \mid \alpha \in \mathcal{O}(\beta)\}$.

$\alpha < \beta \Rightarrow \omega^\alpha + \omega^\beta = \omega^\beta$.

$\omega^0 = 1$.

$\omega^1 = \omega$.

The following result known as the *Cantor Normal Form* for the (countable) ordinals is fundamental.

Proposition 6.20 (Cantor Normal Form) For every ordinal $\alpha \in \mathcal{O}$, if $\alpha > 0$ then there are unique ordinals $\alpha_1 \geq \ldots \geq \alpha_n$, $n \geq 1$, such that

$$\alpha = \omega^{\alpha_1} + \cdots + \omega^{\alpha_n}.$$

Proof. First, we show the existence of the representation. We proceed by transfinite induction. If α is an additive principal ordinal, then $\alpha = \omega^{\alpha_1}$ for some α_1 since $\gamma \mapsto \omega^\gamma$ is the ordering function of the additive principal ordinals. Otherwise, there is some $\delta < \alpha$ such that $\delta + \alpha \neq \alpha$. Then, since $\alpha \leq \delta + \alpha$ (by proposition 6.9), $\delta > 0$ and $\delta + \alpha > \alpha$. Since $\delta < \alpha$, there is some $\eta > 0$ such that $\alpha = \delta + \eta$. We must have $\eta < \alpha$, since otherwise, by right monotonicity, we would have $\delta + \alpha \leq \delta + \eta = \alpha$, contradicting $\delta + \alpha > \alpha$. Hence, $\alpha = \delta + \eta$, with $0 < \delta, \eta < \alpha$. By the induction hypothesis, $\delta = \omega^{\alpha_1} + \cdots + \omega^{\alpha_m}$ and $\eta = \omega^{\beta_1} + \cdots + \omega^{\beta_n}$, for some ordinals such that $\alpha_1 \geq \ldots \geq \alpha_m$ and $\beta_1 \geq \ldots \geq \beta_n$. If we had $\alpha_i < \beta_1$ for all i, $1 \leq i \leq m$, then we would have $\delta + \eta = \eta$ (using the fact that for additive principal ordinals, if $\alpha < \beta$, then $\omega^\alpha + \omega^\beta = \omega^\beta$), that is, $\alpha = \eta$, contradicting the fact that $\eta < \alpha$. Hence, there is a largest k, $1 \leq k \leq m$ such that $\alpha_k \geq \beta_1$. Consequently, $\alpha_1 \geq \ldots \geq \alpha_k \geq \beta_1 \geq \ldots \geq \beta_n$, and since $\omega^{\alpha_j} + \omega^{\beta_1} = \omega^{\beta_1}$ for $k+1 \leq j \leq m$, we have

$$\begin{aligned}\alpha &= \delta + \eta \\ &= \omega^{\alpha_1} + \cdots + \omega^{\alpha_k} + \omega^{\alpha_{k+1}} + \cdots + \omega^{\alpha_m} + \omega^{\beta_1} + \cdots + \omega^{\beta_n} \\ &= \omega^{\alpha_1} + \cdots + \omega^{\alpha_k} + \omega^{\beta_1} + \cdots + \omega^{\beta_n}.\end{aligned}$$

Assume $\alpha = \omega^{\alpha_1} + \cdots + \omega^{\alpha_m} = \omega^{\beta_1} + \cdots + \omega^{\beta_n}$. Uniqueness is shown by induction on m. Note that $\alpha + \omega^{\alpha_1'} = \omega^{\alpha_1'}$, which implies that $\alpha < \omega^{\alpha_1'}$ (by right strict monotonicity, since $\omega^{\alpha_1'} > 0$), and similarly, $\alpha < \omega^{\beta_1'}$. If we had $\beta_1' \leq \alpha_1$, we would have $\omega^{\beta_1'} \leq \omega^{\alpha_1} \leq \alpha$, contradicting the fact that $\alpha < \omega^{\beta_1'}$. Hence, $\alpha_1 < \beta_1'$. Similarly, we have $\beta_1 < \alpha_1'$. But then, $\alpha_1 \leq \beta_1$ and $\beta_1 \leq \alpha_1$, and therefore, $\alpha_1 = \beta_1$. Hence, either $m = n = 1$, or $m, n > 1$ and $\omega^{\alpha_2} + \cdots + \omega^{\alpha_m} = \omega^{\beta_2} + \cdots + \omega^{\beta_n}$. We conclude using the induction hypothesis. □

As we shall see in the next section, there are ordinals such that $\omega^\alpha = \alpha$, and so, we cannot ensure that $\alpha_i < \alpha$. However, if $n > 1$, by right strict monotonicity of $+$, it is true that $\omega^{\alpha_i} < \alpha$, $1 \leq i \leq n$. We are now ready to define some normal functions that will lead us to the definition of Γ_0.

6.5 α-Critical Ordinals

For each $\alpha \in \mathcal{O}$, we shall define a subset $Cr(\alpha) \subseteq \mathcal{O}$ and its ordering function φ_α inductively as follows.

Definition 6.21 For each $\alpha \in \mathcal{O}$, the set $Cr(\alpha) \subseteq \mathcal{O}$ and its ordering function $\varphi_\alpha : A_\alpha \to Cr(\alpha)$ are defined inductively as follows.

(1) $Cr(0)$ = the set of additive principal ordinals, $A_0 = \mathcal{O}$, and for every $\alpha \in \mathcal{O}$, $\varphi_0(\alpha) = \omega^\alpha$, the ordering function of $Cr(0)$.

(2) $Cr(\alpha') = \{\eta \in A_\alpha \mid \varphi_\alpha(\eta) = \eta\}$, the set of fixed points of φ_α, and $\varphi_{\alpha'} : A_{\alpha'} \to Cr(\alpha')$ is the ordering function of $Cr(\alpha')$.

(3) For every limit ordinal $\beta \in \mathcal{O}$,

$$Cr(\beta) = \{\eta \in \bigcap_{\alpha<\beta} A_\alpha \mid \forall \alpha < \beta,\ \varphi_\alpha(\eta) = \eta\},$$

and $\varphi_\beta : A_\beta \to Cr(\beta)$ is the ordering function of $Cr(\beta)$.

The elements of the set $Cr(\alpha)$ are called *α-critical ordinals*. The following proposition shows that for $\alpha > 0$ the α-critical ordinals are the common fixed points of the normal functions φ_β, for all $\beta < \alpha$.

Proposition 6.22 For all $\alpha, \eta \in \mathcal{O}$, if $\alpha = 0$ then $\eta \in Cr(0)$ iff η is additive principal, else $\eta \in Cr(\alpha)$ iff $\eta \in \bigcap_{\beta<\alpha} A_\beta$ and $\varphi_\beta(\eta) = \eta$ for all $\beta < \alpha$.

Proof. We proceed by transfinite induction. The case $\alpha = 0$ is clear since $Cr(0)$ is defined as the set of additive principal ordinals. If α is a successor ordinal, there is some β such that $\alpha = \beta'$. By the induction hypothesis, $\eta \in Cr(\beta)$ iff $\eta \in \bigcap_{\gamma<\beta} A_\gamma$ and $\varphi_\gamma(\eta) = \eta$ for all $\gamma < \beta$. By the definition of $Cr(\beta')$, $\eta \in Cr(\beta') = Cr(\alpha)$ iff $\eta \in A_\beta$ and $\varphi_\beta(\eta) = \eta$. Hence, since $\alpha = \beta'$, $\eta \in Cr(\alpha)$ iff $\eta \in \bigcap_{\gamma<\alpha} A_\gamma$ and $\varphi_\gamma(\eta) = \eta$ for all $\gamma < \alpha$. If α is a limit ordinal, the property to be shown is clause (3) of definition 6.21. □

The following important result holds.

Proposition 6.23 Each set $Cr(\alpha)$ is closed and unbounded.

Proof. We show by transfinite induction that $Cr(\alpha)$ is closed and unbounded and that $A_\alpha = \mathcal{O}$.

Proof of closure. For $\alpha = 0$ this follows from the fact the the set of additive principal ordinals is closed. Assume $\alpha > 0$, and let $M \subseteq Cr(\alpha)$ be a nonempty countable subset of $Cr(\alpha)$. By the induction hypothesis, for every $\beta < \alpha$, $Cr(\beta)$ is closed and $A_\beta = \mathcal{O}$. Hence, by proposition 6.13, φ_β is continuous. Hence, $\varphi_\beta(\bigsqcup M) = \bigsqcup M$ for all $\beta < \alpha$. By proposition 6.22, since we also have $A_\beta = \mathcal{O}$ for all $\beta < \alpha$, this implies that $\bigsqcup M \in Cr(\alpha)$. Hence, $Cr(\alpha)$ is closed.

Proof of Unboundedness. For $\alpha = 0$, this follows from the fact that the set of additive principal ordinals in unbounded and that $A_0 = \mathcal{O}$. Assume $\alpha > 0$. Given any ordinal β, let $\gamma_0 = \beta'$, $\gamma_{n+1} = \bigsqcup\{\varphi_\eta(\gamma_n) \mid \eta < \alpha\}$, $M = \{\gamma_n \mid n \in \mathbf{N}\}$, and $\gamma = \bigsqcup M$. By the induction hypothesis, for every $\delta < \alpha$, $Cr(\delta)$ is unbounded, and so γ_n is well defined for all $n \geq 0$. We have $\beta < \gamma_0 \leq \gamma$. For every $\delta < \alpha$, we have $\varphi_\delta(\gamma_n) \leq \gamma_{n+1} \leq \gamma$, and so

$\bigsqcup\{\varphi_\delta(\gamma_n) \mid \gamma_n \in M\} \leq \gamma$. By the induction hypothesis, for every $\delta < \alpha$, $Cr(\delta)$ is closed and unbounded and $A_\delta = \mathcal{O}$. Hence, φ_δ is continuous and

$$\varphi_\delta(\bigsqcup M) = \bigsqcup\{\varphi_\delta(\gamma_n) \mid \gamma_n \in M\}.$$

Hence, $\varphi_\delta(\gamma) \leq \gamma$. By proposition 6.9, we also have $\gamma \leq \varphi_\delta(\gamma)$. Hence, $\gamma = \varphi_\delta(\gamma)$ for all $\delta < \alpha$. By proposition 6.22, we have $\gamma \in Cr(\alpha)$, and γ is an α-critical ordinal $> \beta$. Hence $Cr(\alpha)$ is unbounded, and so $A_\alpha = \mathcal{O}$. □

Proposition 6.23 has the following corollary.

Proposition 6.24 For every $\alpha \in \mathcal{O}$, $A_\alpha = \mathcal{O}$ and φ_α is a normal function.

In view of proposition 6.24, since every function φ_α has domain $\mathcal{O}$, we can define the function $\varphi : \mathcal{O} \times \mathcal{O} \to \mathcal{O}$ such that $\varphi(\alpha, \beta) = \varphi_\alpha(\beta)$ for all $\alpha, \beta \in \mathcal{O}$. From definition 6.21 and proposition 6.24, we have the following useful properties.

Proposition 6.25 (1) $\eta \in Cr(\alpha')$ iff $\varphi(\alpha, \eta) = \eta$.

(2) For a limit ordinal β, $Cr(\beta) = \bigcap_{\alpha<\beta} Cr(\alpha)$.

Proposition 6.26 (1) If $\alpha < \beta$ then $Cr(\beta) \subseteq Cr(\alpha)$.

(2) Every ordinal $\varphi(\alpha, \beta)$ is an additive principal ordinal.

(3) $\varphi(0, \beta) = \omega^\beta$.

An ordinal α such that $\alpha \in Cr(\alpha)$ is particularly interesting. Actually, it is by no means obvious that such ordinals exist, but they do, and Γ_0 is the smallest. We shall consider this property in more detail.

It is interesting to see what are the elements of $Cr(1)$. By the definition, an ordinal α is in $Cr(1)$ iff $\omega^\alpha = \alpha$. Such ordinals are called *epsilon ordinals*, because their ordering function is usually denoted by ϵ. The least element of $Cr(1)$ is ϵ_0. It can be shown that ϵ_0 is the least upper bound of the set

$$\{\omega, \omega^\omega, \omega^{\omega^\omega}, \ldots, \omega^{\omega^{\cdot^{\cdot^{\cdot^\omega}}}}, \ldots\}.$$

This is already a rather impressive ordinal. What are the elements of $Cr(2)$? Well, denoting the ordering function of $Cr(1)$ by ϵ, $\alpha \in Cr(2)$ iff $\epsilon_\alpha = \alpha$. We claim that the smallest of these ordinals is greater than

$$\epsilon_0, \epsilon_1, \ldots, \epsilon_\omega, \ldots, \epsilon_{\epsilon_0}, \ldots, \epsilon_{\epsilon_1}, \ldots, \epsilon_{\epsilon_{\epsilon_0}}, \ldots$$

Amazingly, the ordinal Γ_0 *dwarfs* the ordinals just mentioned, and many more!

The following proposition gives a rather explicit characterization of $\varphi_{\alpha'}$ in terms of fixed points. It also shows that the first element of $Cr(\alpha')$ is farther down than the first element of $Cr(\alpha)$ on the ordinal line (in fact, much farther down).

Proposition 6.27 For each $\alpha, \beta \in \mathcal{O}$, let $\varphi_\alpha^0(\beta) = \beta$, and $\varphi_\alpha^{n+1}(\beta) = \varphi_\alpha(\varphi_\alpha^n(\beta))$ for every $n \geq 0$. Then, we have

$$\varphi_{\alpha'}(0) = \bigsqcup_{n\geq 0} \varphi_\alpha^n(0),$$
$$\varphi_{\alpha'}(\beta') = \bigsqcup_{n\geq 0} \varphi_\alpha^n(\varphi_{\alpha'}(\beta) + 1),$$
$$\varphi_{\alpha'}(\beta) = \bigsqcup_{\gamma<\beta} \varphi_{\alpha'}(\gamma),$$

for a limit ordinal β. Furthermore, $\varphi_\alpha(0) < \varphi_{\alpha'}(0)$ for all $\alpha \in \mathcal{O}$.

Proof. Since φ_α is a normal function, by proposition 6.15, $\bigsqcup_{n\geq 0} \varphi_\alpha^n(0)$ is the least fixed point of φ_α, and for every $\beta \in \mathcal{O}$, $\bigsqcup_{n\geq 0} \varphi_\alpha^n(\varphi_{\alpha'}(\beta) + 1)$ is the least fixed point of φ_α that is $> \varphi_{\alpha'}(\beta)$. Since $\varphi_{\alpha'}$ enumerates the fixed points of φ_α, $\varphi_{\alpha'}(\beta') = \bigsqcup_{n\geq 0} \varphi_\alpha^n(\varphi_{\alpha'}(\beta) + 1)$.

Assume that β is a limit ordinal. From the proof of proposition 6.4, we know that $\beta = \bigsqcup\{\gamma \mid \gamma < \beta\}$. Since $\varphi_{\alpha'}$ is continuous, we have

$$\varphi_{\alpha'}(\beta) = \varphi_{\alpha'}(\bigsqcup\{\gamma \mid \gamma < \beta\}) = \bigsqcup_{\gamma<\beta} \varphi_{\alpha'}(\gamma).$$

Since $0 < \varphi_\alpha(0)$, it is easily shown that $\varphi_\alpha^n(0) < \varphi_\alpha^{n+1}(0)$ for all $n \geq 0$ (using induction and the fact that φ_α is strictly monotonic), and so, $\varphi_\alpha^n(0) < \varphi_{\alpha'}(0)$. Since $\varphi_\alpha^1(0) = \varphi_\alpha(0)$, the first element of $Cr(\alpha)$, we have $\varphi_\alpha(0) < \varphi_{\alpha'}(0)$. □

Proposition 6.27 justifies the claim we made about ϵ_0, and also shows that the first element of $Cr(2)$ is the least upper bound of the set

$$\{\epsilon_0, \epsilon_{\epsilon_0}, \epsilon_{\epsilon_{\epsilon_0}}, \ldots, \epsilon_{\epsilon_{\cdot^{\cdot^{\cdot_{\epsilon_0}}}}}, \ldots\}$$

It is hard to conceive what this limit is! Of course, things get worse when we look at the first element of $Cr(3)$, not to mention the notational difficulties involved. Can you imagine what the first element of $Cr(\epsilon_0)$ is? Well, Γ_0 is farther away on the ordinal line!

The following proposition characterizes the order relationship between $\varphi(\alpha_1, \beta_1)$ and $\varphi(\alpha_2, \beta_2)$.

Proposition 6.28 (i) $\varphi(\alpha_1, \beta_1) = \varphi(\alpha_2, \beta_2)$ iff either

(1) $\alpha_1 < \alpha_2$ and $\beta_1 = \varphi(\alpha_2, \beta_2)$, or

(2) $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$, or

(3) $\alpha_2 < \alpha_1$ and $\varphi(\alpha_1, \beta_1) = \beta_2$.

(ii) $\varphi(\alpha_1, \beta_1) < \varphi(\alpha_2, \beta_2)$ iff either

(1) $\alpha_1 < \alpha_2$ and $\beta_1 < \varphi(\alpha_2, \beta_2)$, or

(2) $\alpha_1 = \alpha_2$ and $\beta_1 < \beta_2$, or

(3) $\alpha_2 < \alpha_1$ and $\varphi(\alpha_1, \beta_1) < \beta_2$.

Proof (sketch). We sketch the proof of (ii). By the definition of φ, $\varphi(\alpha_2, \beta_2) \in Cr(\alpha_2)$. If $\alpha_1 < \alpha_2$, by proposition 6.22, $\varphi(\alpha_2, \beta_2)$ is a fixed point of φ_{α_1}, and so,

$$\varphi(\alpha_1, \varphi(\alpha_2, \beta_2)) = \varphi(\alpha_2, \beta_2).$$

Since φ_{α_1} is strictly monotonic, $\varphi(\alpha_1, \beta_1) < \varphi(\alpha_1, \varphi(\alpha_2, \beta_2))$ iff $\beta_1 < \varphi(\alpha_2, \beta_2)$. The case where $\alpha_2 < \alpha_1$ is similar. For $\alpha_1 = \alpha_2$, the assertion follows from the fact that φ_{α_1} is strictly monotonic. □

Using proposition 6.9, since each function φ_α is an ordering function, we have the following useful property.

Proposition 6.29 For all $\alpha, \beta \in \mathcal{O}$, $\beta \leq \varphi(\alpha, \beta)$.

By proposition 6.28 and 6.29, we also have the following.

Corollary 6.30 For all $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathcal{O}$, if $\alpha_1 \leq \alpha_2$ and $\beta_1 \leq \beta_2$, then $\varphi(\alpha_1, \beta_1) \leq \varphi(\alpha_2, \beta_2)$.

The following can be shown by transfinite induction.

Proposition 6.31 (i) For every $\alpha \in \mathcal{O}$, $\alpha \leq \varphi(\alpha, 0)$. Furthermore, if $\beta \in Cr(\alpha)$, then $\alpha \leq \beta$.

(ii) If $\alpha \leq \beta$, then $\varphi(\alpha, \beta) \leq \varphi(\beta, \alpha)$.

Proof. We show $\alpha \leq \varphi(\alpha, 0)$ by transfinite induction. This is clear for $\alpha = 0$. If $\alpha > 0$, for every $\beta < \alpha$, by strict monotonicity and proposition 6.22, $\varphi(\beta, 0) < \varphi(\beta, \varphi(\alpha, 0)) = \varphi(\alpha, 0)$, since $\varphi(\alpha, 0) > 0$ is a fixed point of φ_β. By the induction hypothesis, we have $\beta \leq \varphi(\beta, 0)$, and so $\beta < \varphi(\alpha, 0)$ for all $\beta < \alpha$. By proposition 6.4, this implies that $\alpha \leq \varphi(\alpha, 0)$.

$\beta \in Cr(\alpha)$ iff $\beta = \varphi(\alpha, \eta)$ for some η, and since $\alpha \leq \varphi(\alpha, 0)$, by monotonicity, we have $\alpha \leq \varphi(\alpha, 0) \leq \varphi(\alpha, \eta) = \beta$.

Assume $\alpha \leq \beta$. Since $\beta \leq \varphi(\beta, 0)$, we also have $\beta \leq \varphi(\beta, \alpha)$. By proposition 6.28, $\varphi(\alpha, \beta) \leq \varphi(\beta, \alpha)$, since $\alpha \leq \beta$ and $\beta \leq \varphi(\beta, \alpha)$. □

Another key result is the following.

Proposition 6.32 For every additive principal ordinal γ, there exist unique $\alpha, \beta \in \mathcal{O}$ such that, $\alpha \leq \gamma$, $\beta < \gamma$, and $\gamma = \varphi(\alpha, \beta)$.

Proof. Recall that an additive principal ordinal is not equal to 0. By proposition 6.31, $\gamma \leq \varphi(\gamma, 0)$. Since $0 < \gamma$, by strict monotonicity of φ_γ, $\varphi(\gamma, 0) < \varphi(\gamma, \gamma)$, and so $\gamma < \varphi(\gamma, \gamma)$. Since $\mathcal{O}$ is well-ordered, there is a least ordinal $\alpha \leq \gamma$ such that $\gamma < \varphi(\alpha, \gamma)$. If $\alpha \neq 0$, the minimality of α implies that $\varphi(\eta, \gamma) = \gamma$ for all $\eta < \alpha$, and by proposition 6.22, $\gamma \in Cr(\alpha)$. If $\alpha = 0$, since γ is an additive principal ordinal, by the definition of $Cr(0)$, $\alpha \in Cr(0)$. Hence, $\gamma \in Cr(\alpha)$. Hence, there is some β such that $\gamma = \varphi(\alpha, \beta)$. Since $\gamma < \varphi(\alpha, \gamma)$, by strict monotonicity of φ_α, we must have $\beta < \gamma$.

It remains to prove the uniqueness of α and β. If $\beta_1 < \gamma$, $\beta_2 < \gamma$, and $\gamma = \varphi(\alpha_1, \beta_1) = \varphi(\alpha_2, \beta_2)$, by proposition 6.28, we must have $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$. □

Observe that the proof does not show that $\alpha < \gamma$, and indeed, this is not necessarily true. Also, for an ordinal γ, $\gamma = \varphi(\gamma, \beta)$ holds for some β iff $\gamma \in Cr(\gamma)$. Such ordinals exist in abundance, as we shall prove next.

Definition 6.33 An ordinal $\alpha \in \mathcal{O}$ is a *strongly critical ordinal* iff $\alpha \in Cr(\alpha)$.

Proposition 6.34 An ordinal α is strongly critical iff $\varphi(\alpha, 0) = \alpha$.

Proof. If $\alpha \in Cr(\alpha)$, there is some β such that $\alpha = \varphi(\alpha, \beta)$. By proposition 6.31, we have $\alpha \leq \varphi(\alpha, 0)$, and by strict monotonicity of φ_α, we have $\beta = 0$. Conversely, it is obvious that $\varphi(\alpha, 0) = \alpha$ implies $\alpha \in Cr(\alpha)$. □

Let $\psi : \mathcal{O} \to \mathcal{O}$ be the function defined such that $\psi(\alpha) = \varphi(\alpha, 0)$ for all $\alpha \in \mathcal{O}$. We shall prove that ψ is strictly monotonic and continuous. As a consequence, ψ is a normal function for the set $\{\varphi(\alpha, 0) \mid \alpha \in \mathcal{O}\}$.

Proposition 6.35 The function ψ (also denoted by $\varphi(-, 0)$) defined such that $\psi(\alpha) = \varphi(\alpha, 0)$ for all $\alpha \in \mathcal{O}$ is strictly monotonic and continuous.

Proof. First, we prove the following claim.

Claim: ψ satisfies the following properties:

$$\begin{aligned}\psi(0) &= \varphi(0,0),\\ \psi(\beta') &= \bigsqcup_{n\geq 0} \varphi_\beta^n(\psi(\beta)),\\ \psi(\beta) &= \bigsqcup_{\delta<\beta} \psi(\delta),\end{aligned}$$

for a limit ordinal β.

Proof of claim. By definition, $\psi(0) = \varphi(0,0)$, and the second identity follows from proposition 6.15, since $\varphi_\beta^1(0) = \varphi(\beta,0) = \psi(\beta)$, which implies that $\varphi_\beta^n(\psi(\beta)) = \varphi_\beta^{n+1}(0)$ for all $n \geq 0$. By proposition 6.22, $\psi(\beta) = \varphi(\beta,0) = \eta_0$, where η_0 is the least ordinal such that $\varphi(\gamma,\eta) = \eta$ for all $\gamma < \beta$. For every $\gamma < \beta$, since φ_γ is continuous,

$$\begin{aligned}\varphi(\gamma, \bigsqcup_{\delta<\beta} \psi(\delta)) &= \bigsqcup_{\delta<\beta} \varphi(\gamma,\psi(\delta))\\ &= \bigsqcup_{\delta<\beta} \varphi(\gamma,\varphi(\delta,0)).\end{aligned}$$

For $\delta > \gamma$, we have $\varphi(\gamma,\varphi(\delta,0)) = \varphi(\delta,0) = \psi(\delta)$, and since φ is monotonic in both arguments,

$$\bigsqcup_{\delta<\beta} \varphi(\gamma,\varphi(\delta,0)) = \bigsqcup_{\delta<\beta} \psi(\delta).$$

Hence,

$$\varphi(\gamma, \bigsqcup_{\delta<\beta} \psi(\delta)) = \bigsqcup_{\delta<\beta} \psi(\delta),$$

for all $\gamma < \beta$, which shows that $\eta_0 \leq \bigsqcup_{\delta<\beta} \psi(\delta)$ (because η_0 is the least such common fixed point). On the other hand, $\psi(\delta) = \varphi(\delta,0) \leq \varphi(\delta,\eta_0) = \eta_0$ for all $\delta < \beta$. Hence, $\bigsqcup_{\delta<\beta} \psi(\delta) \leq \eta_0$. But then, $\bigsqcup_{\delta<\beta} \psi(\delta) = \eta_0 = \psi(\beta)$. □

We can now show that ψ is continuous. Let M be a nonempty countable subset of $\mathcal{O}$, and let $\beta = \bigsqcup M$. The case $\beta = 0$ is trivial. If $\beta = \alpha'$ for some α, we must have $\beta \in M$, since otherwise β would not be the least upper bound of M (either $\gamma \leq \alpha$ for all $\gamma \in M$, or $\gamma > \alpha$ for some $\gamma \in M$, a contradiction in either case). But then, $\psi(\bigsqcup M) = \psi(\beta) = \bigsqcup_{\alpha\in M} \psi(\alpha)$, since ψ is monotonic. If $\beta = \bigsqcup M$ is a limit ordinal, then $\beta = \bigsqcup M = \bigsqcup\{\delta \mid \delta < \beta\}$. Hence, for every $\alpha \in M$, there is some $\delta < \beta$ such that $\alpha < \delta$, and conversely, for every $\delta < \beta$, there is some $\alpha \in M$ such that $\delta < \alpha$. By monotonicity of ψ, this implies that

$$\bigsqcup_{\alpha\in M} \psi(\alpha) = \bigsqcup_{\delta<\beta} \psi(\delta).$$

By the claim,

$$\psi(\bigsqcup M) = \psi(\beta) = \bigsqcup_{\delta<\beta} \psi(\delta),$$

and therefore,

$$\psi(\bigsqcup M) = \bigsqcup_{\alpha\in M} \psi(\alpha),$$

showing that ψ is continuous.

Finally, we show that ψ is strictly monotonic. Since φ is monotonic in both arguments, $\psi = \varphi(-,0)$ is monotonic. Assume $\alpha < \beta$. Then $\alpha < \alpha' \leq \beta$ and by proposition 6.27, $\psi(\alpha) < \psi(\alpha') \leq \psi(\beta)$. □

Proposition 6.35 implies that there are plenty of strongly critical ordinals.

Proposition 6.36 The set of strongly critical ordinals is closed and unbounded.

Proof. First, we prove unboundedness. Since $\psi = \varphi(-,0)$ is a normal function, by proposition 6.22, for any arbitrary ordinal α, ψ has a least fixed point $> \alpha$. Since such fixed points are strongly critical ordinal, the set of strongly critical ordinals is unbounded.

Next, we prove that the set of strongly critical ordinals is closed. Let M be a nonempty countable set of strongly critical ordinals. For each $\alpha \in M$, we have $\psi(\alpha,0) = \alpha$. Hence, $\psi(M) = M$. Since $\psi = \varphi(-,0)$ is continuous, we have $\psi(\bigsqcup M) = \bigsqcup \psi(M) = \bigsqcup M$. This shows that $\bigsqcup M$ is a strongly critical ordinal, and therefore, the set of strongly critical ordinals is closed. □

From proposition 6.36, the ordering function of the set of strongly critical ordinals is a normal function. This function is denoted by Γ, and $\Gamma(0)$, also denoted Γ_0, is the least strongly critical ordinal. Γ_0 is the least ordinal such that $\varphi(\alpha,0) = \alpha$. The following proposition shows that $\mathcal{O}(\Gamma_0)$ is closed under $+$ and φ.

Proposition 6.37 For all $\alpha, \beta \in \mathcal{O}$, if $\alpha, \beta < \Gamma_0$, then $\alpha + \beta < \Gamma_0$, and $\varphi(\alpha,\beta) < \Gamma_0$.

Proof (sketch). Since Γ_0 is an additive principal ordinal, closure under $+$ is clear. Let $\gamma_0 = 0$, $\gamma_{n+1} = \varphi(\gamma_n, 0)$, $U = \{\gamma_n \mid n \in \mathbf{N}\}$, and $\gamma = \bigsqcup U$. By proposition 6.15, we have $\gamma = \Gamma_0$. Now, if $\alpha, \beta < \Gamma_0$, since $\Gamma_0 = \bigsqcup U$, there is some γ_n such that $\alpha, \beta < \gamma_n$. By proposition 6.28, we have $\varphi(\alpha,\beta) < \varphi(\gamma_n, 0)$, because $\beta < \gamma_n \leq \varphi(\gamma_n, 0)$. Hence, $\varphi(\alpha,\beta) < \gamma_{n+1} \leq \Gamma_0$. □

Proposition 6.37 shows that Γ_0 cannot be obtained from strictly smaller ordinals in terms of the function $+$ and the powerful functions φ_α, $\alpha < \Gamma_0$. As Smoryński puts it in one of his articles [50],

"Γ_0 is the first countable ordinal which cannot be described without reference (if only oblique) to the uncountable."

Indeed, referring to Γ_0 as the *least* ordinal α satisfying $\alpha = \varphi(\alpha, 0)$ is indirect and somewhat circular – the word "least" involves reference to all ordinals, including Γ_0. One could claim that the definition of Γ_0 as $\bigsqcup\{\gamma_n \mid n \in \mathbf{N}\}$, as in proposition 6.37, is "constructive", and does not refer to the uncountable, but this is erroneous, although the error is more subtle. Indeed, the construction of the function $\varphi(-, 0)$ is actually an iteration of the *functional* taking us from $\varphi(\alpha, -)$ to $\varphi(\alpha', -)$, and therefore, presupposes as domain of this functional a class of functions on ordinals and thus (on close examination) the uncountable. As logicians say, the definition of the ordinal Γ_0 is *impredicative*.

7 A Glimpse at Veblen Hierarchies

What have we accomplished in section 6.5? If one examines carefully the proofs of propositions 6.23, 6.24, 6.27, 6.28, 6.31, 6.35, and definition 6.21, one discovers that the conditions that make everything go through are the fact that $\alpha \mapsto \omega^\alpha$ is a normal function φ such that $0 < \varphi(0)$. This suggests the following generalization.

Definition 7.1 Given *any* normal function φ such that $0 < \varphi(0)$, mimicking definition 6.21, we define the hierarchy $\{\varphi^0_\alpha\}_{\alpha \in \mathcal{O}}$ of functions such that,

- $\varphi^0_0 = \varphi$, and for every $\alpha > 0$,
- φ^0_α enumerates the set $\{\eta \mid \varphi^0_\beta(\eta) = \eta, \text{ for all } \beta < \alpha\}$ of common fixed points of the functions φ^0_β for all $\beta < \alpha$.

We have what is called a *Veblen hierarchy* (a concept due to Veblen [53]), and according to our previous remark, the following properties hold.

Theorem 7.2 (Veblen Hierarchy theorem) Denoting each function φ^0_α as $\varphi^0(\alpha, -)$, each $\varphi^0(\alpha, -)$ is a normal function, and the function $\varphi^0(-, 0) : \alpha \mapsto \varphi^0(\alpha, 0)$ is also a normal function such that $0 < \varphi^0(0, 0)$.

But since $\varphi^0(-, 0)$ satisfies the conditions for building a Veblen hierarchy, we can iterate the process just described in definition 7.1. For this, following Larry Miller [34], it is convenient to define an operator Δ_1 on normal functions, the *1-diagonalization operator*, defined as follows.

Given a normal function φ such that $0 < \varphi(0)$, $\Delta_1(\varphi)$ is the normal function enumerating the fixed points of $\varphi^0(-, 0)$.

Note that in a *single step*, Δ_1 performs the Ω iterations producing the Veblen hierarchy $\{\varphi^0_\alpha\}_{\alpha < \Omega}$! (where Ω denotes the first uncountable ordinal, i.e., the order type of $\mathcal{O}$). Using the operator Δ_1, we can define a sequence $\{\varphi^1_\beta\}_{\beta < \Omega}$ of normal functions, and so, a sequence of Veblen hierarchies – or a doubly indexed sequence of normal functions – $\{\varphi^1_\beta(\gamma, -)\}_{\beta, \gamma < \Omega}$ defined as follows:

- $\varphi^1_0 = \varphi$,
- $\varphi^1_{\beta'} = \Delta_1(\varphi^1_\beta)$, and
- φ^1_β is the normal function enumerating $\bigcap_{\gamma < \beta} range(\varphi^1_\gamma)$, for a limit ordinal β.

But $\beta \mapsto \varphi^1_\beta(0)$ (also denoted $\varphi^1(-, 0)$) is also a normal function such that $0 < \varphi^1_0(0)$. Hence, we can define an operator Δ_2 enumerating the fixed points of $\beta \mapsto \varphi^1_\beta(0)$, and build

a hierarchy. But we can iterate the operator Δ into the transfinite! This leads to the following definition.

Definition 7.3 Given a normal function φ such that $0 < \varphi(0)$, we define by simultaneous induction the Ω-indexed sequence $\{\Delta_\alpha\}_{\alpha<\Omega}$ of diagonalization operators and the doubly Ω-indexed sequence $\{\varphi^\alpha_\beta\}_{\alpha,\beta<\Omega}$ of normal functions as follows.

- $\Delta_0(\psi)$ enumerates the fixed points of the normal function ψ;
- $\Delta_{\alpha'}(\varphi) = \Delta_0(\varphi^\alpha(-,0))$ enumerates the fixed points of $\varphi^\alpha(-,0) : \beta \mapsto \varphi^\alpha_\beta(0)$;
- $\Delta_\alpha(\varphi)$ enumerates $\bigcap_{\gamma<\alpha} range(\Delta_\gamma(\varphi))$, for a limit ordinal α;
- $\varphi^\alpha_0 = \varphi$;
- $\varphi^\alpha_{\beta'} = \Delta_\alpha(\varphi^\alpha_\beta)$;
- φ^α_β enumerates $\bigcap_{\gamma<\beta} range(\varphi^\alpha_\gamma)$, for a limit ordinal β.

It is convenient to keep track of the diagonalization level (the index α) and the number of iterations of diagonalizations of level α (the index β) by using indices beyond Ω. Indeed, using the families $\{\varphi^\alpha_\beta\}_{\alpha,\beta<\Omega}$ and the representation of the ordinals in base Ω, it is possible to extend our original Ω-indexed hierarchy $\{\varphi(\beta,-)\}_{\beta<\Omega}$ (dropping the superscript 0 in φ^0) to an Ω^Ω-indexed hierarchy $\{\varphi(\delta,-)\}_{\delta<\Omega^\Omega}$. Let us first consider the simple case where $\alpha = 1$.

Using the fact that every ordinal $\delta < \Omega^2$ is uniquely expressed as $\delta = \Omega\beta_1 + \beta_2$ for some ordinals $\beta_1, \beta_2 < \Omega$, we can extend the Ω-indexed hierarchy $\{\varphi(\beta,-)\}_{\beta<\Omega}$ to an Ω^2-indexed hierarchy $\{\varphi(\delta,-)\}_{\delta<\Omega^2}$ as follows. For any $\beta_1, \beta_2 < \Omega$, we let

$$\varphi(\Omega\beta_1 + \beta_2, -) = (\varphi^1_{\beta_1})^0_{\beta_2}.$$

With this convention applied to the function $\omega(-) : \alpha \mapsto \omega^\alpha$ and the Ω^2-indexed sequence $\{\omega(\delta,-)\}_{\delta<\Omega^2}$, note that $\omega^1_1 = \Delta_1(\omega(-)) = \Delta_0(\omega^0(-,0))$ is denoted by $\omega(\Omega,-)$, and $\omega(\Omega,0) = \Gamma_0$ denotes the least fixed point of $\omega^0(-,0)$. Similarly, $\omega^2_1 = \Delta_2(\omega(-)) = \Delta_0(\omega^1(-,0))$ is denoted by $\omega(\Omega^2,-)$, and $\omega(\Omega^2,0)$ denotes the least fixed point of $\omega^1(-,0)$.

In general, since every ordinal $\delta < \Omega^\Omega$ is uniquely expressed as

$$\delta = \Omega^{\alpha_1}\beta_1 + \cdots + \Omega^{\alpha_n}\beta_n$$

for some ordinals $\alpha_n < \ldots < \alpha_1 < \Omega$ and $\beta_1, \ldots, \beta_n < \Omega$, we can regard the multiply Ω-indexed sequence

$$\{(\cdots(\varphi^{\alpha_1}_{\beta_1})\cdots)^{\alpha_n}_{\beta_n})\}_{\alpha_n<\ldots<\alpha_1<\Omega,\beta_1,\ldots,\beta_n<\Omega}$$

as an Ω^Ω-indexed sequence $\{\varphi(\delta, -)\}_{\delta<\Omega^\Omega}$, if we put

$$\varphi(\Omega^{\alpha_1}\beta_1 + \cdots + \Omega^{\alpha_n}\beta_n, -) = (\cdots(\varphi_{\beta_1}^{\alpha_1})\cdots)_{\beta_n}^{\alpha_n}).$$

Hence, a constructive ordinal notation system for the ordinals less than $\varphi(\Omega^\Omega, 0)$, the least fixed point of $\delta \mapsto \varphi(\delta, 0)$ $(\delta < \Omega^\Omega)$, can be obtained using the families

$$\{(\cdots(\varphi_{\beta_1}^{\alpha_1})\cdots)_{\beta_n}^{\alpha_n})\}_{\alpha_n<\ldots<\alpha_n<\Omega,\beta_1,\ldots,\beta_n<\Omega}.$$

It is possible to go farther using Bachmann-Isles hierarchies, but we are already quite dizzy, and refer the reader to Larry Miller's paper [34]. Readers interested in the topic of ordinal notations should consult the very nice expository articles by Crossley and Bridge Kister [5], Miller [34], and Pohlers [42], and for deeper results, Schütte [46] and Pohlers [41].

8 Normal Form For the Ordinals $< \Gamma_0$

One of the most remarkable properties of Γ_0 is that the ordinals less than Γ_0 can be represented in terms of the functions $+$ and φ. First, we need the following lemma.

Lemma 8.1 Given an additive principal ordinal γ, if $\gamma = \varphi(\alpha, \beta)$, with $\alpha \leq \gamma$ and $\beta < \gamma$, then $\alpha < \gamma$ iff γ is not strongly critical.

Proof. By proposition 6.31, we have $\gamma \leq \varphi(\gamma, 0)$. By proposition 6.28, since $\alpha \leq \gamma$ and $\beta < \gamma \leq \varphi(\gamma, 0)$, we have $\gamma = \varphi(\alpha, \beta) < \varphi(\gamma, 0)$ iff $\alpha < \gamma$. By proposition 6.34 and proposition 6.31, γ is not critical iff $\gamma < \varphi(\gamma, 0)$, iff $\alpha < \gamma$ from above. □

We can now prove the fundamental normal form representation theorem for the ordinals less than Γ_0.

Theorem 8.2 For every ordinal α such that $0 < \alpha < \Gamma_0$, there exist unique ordinals $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$, $n \geq 1$, with $\alpha_i, \beta_i < \varphi(\alpha_i, \beta_i) \leq \alpha$, $1 \leq i \leq n$, such that

(1) $\alpha = \varphi(\alpha_1, \beta_1) + \cdots + \varphi(\alpha_n, \beta_n)$, and

(2) $\varphi(\alpha_1, \beta_1) \geq \ldots \geq \varphi(\alpha_n, \beta_n)$.

Proof. Using the Cantor Normal Form for the (countable) ordinals (proposition 6.20), there are unique ordinals $\eta_1 \geq \ldots \geq \eta_n$, $n \geq 1$, such that

$$\alpha = \omega^{\eta_1} + \cdots + \omega^{\eta_n}.$$

Each ordinal ω^{η_i} is an additive principal ordinal, and let $\gamma_i = \omega^{\eta_i}$. By Proposition 6.32, for every additive principal ordinal γ_i, there exist unique $\alpha_i, \beta_i \in \mathcal{O}$ such that, $\alpha_i \leq \gamma_i$, $\beta_i < \gamma_i$,

and $\gamma_i = \varphi(\alpha_i, \beta_i)$. Since for each ordinal γ_i, we have $\gamma_i \leq \alpha < \Gamma_0$, and Γ_0 is the least strongly critical ordinal, by proposition 8.1, $\alpha_i < \gamma_i$. Since $\gamma_i \leq \alpha$, $\alpha_i < \gamma_i$, and $\beta_i < \gamma_i$, we have $\alpha_i < \alpha$ and $\beta_i < \alpha$. Property (2) follows from the fact that $\eta_1 \geq \ldots \geq \eta_n$ implies that $\gamma_1 \geq \ldots \geq \gamma_n$ (since $\gamma_i = \omega^{\eta_i}$). □

We need a few more properties of the ordinals less than Γ_0 before we establish the connection between Γ_0 and Kruskal's theorem.

Lemma 8.3 For all $\alpha, \beta < \Gamma_0$, if $\alpha \leq \beta$, then

$$\alpha \leq \beta \leq \beta + \alpha \leq \varphi(\beta, \alpha),$$

and if $\alpha \leq \beta$ and $\beta < \varphi(\alpha, \beta)$, then

$$\beta + \alpha \leq \varphi(\alpha, \beta) \leq \varphi(\beta, \alpha).$$

Proof. That $\alpha \leq \beta \leq \beta + \alpha$ is easy to show. If $\alpha = 0$, since by proposition 6.31, $\beta \leq \varphi(\beta, 0)$, we have $\beta + 0 = \beta \leq \varphi(\beta, 0)$. If $0 < \alpha = \beta$, we have shown earlier that $\alpha < \varphi(\alpha, \alpha)$ (in the proof of proposition 6.32), and since $\varphi(\alpha, \alpha)$ is an additive principal ordinal, we also have $\alpha + \alpha < \varphi(\alpha, \alpha)$. If $0 < \alpha < \beta$, by proposition 6.29, we have $\beta \leq \varphi(0, \beta)$, and by proposition 6.31, we have $\beta \leq \varphi(\beta, 0)$. By strict monotonicity of φ_β, since $\alpha > 0$, we have $\beta < \varphi(\beta, \alpha)$. Hence, $\alpha < \beta < \varphi(\beta, \alpha)$. By proposition 6.28, $\varphi(0, \beta) < \varphi(\beta, \alpha)$, since $\beta < \varphi(\beta, \alpha)$. Hence,

$$\beta + \alpha \leq \varphi(0, \beta) + \varphi(\beta, \alpha) = \varphi(\beta, \alpha),$$

since $\varphi(0, \beta) < \varphi(\beta, \alpha)$ and $\varphi(\beta, \alpha)$ is an additive principal ordinal.

Now assume $\alpha \leq \beta$ and $\beta < \varphi(\alpha, \beta)$. If $\alpha = 0$, since by proposition 6.29, $\beta \leq \varphi(0, \beta)$, we have $\beta + 0 = \beta \leq \varphi(0, \beta)$. If $0 < \alpha = \beta$, the proof is identical to the proof of the previous case. If $0 < \alpha < \beta$, then by proposition 6.28, $\varphi(0, \beta) < \varphi(\alpha, \beta)$, since $\beta < \varphi(\alpha, \beta)$. We can also show that $\alpha < \varphi(\alpha, \beta)$ as in the previous case (since $\beta > 0$), and we have

$$\beta + \alpha \leq \varphi(0, \beta) + \varphi(\alpha, \beta) = \varphi(\alpha, \beta),$$

since $\varphi(0, \beta) < \varphi(\alpha, \beta)$ and $\varphi(\alpha, \beta)$ is an additive principal ordinal. The fact that $\varphi(\alpha, \beta) \leq \varphi(\beta, \alpha)$ if $\alpha \leq \beta$ was shown in proposition 6.31. □

It should be noted that if $\alpha \leq \beta$, when $\beta = \varphi(\alpha, \beta)$ (which happens when $\beta \in Cr(\alpha')$), the inequality $\beta + \alpha \leq \varphi(\alpha, \beta)$ is *incorrect*. This minor point noted at the very end of Simpson's paper [47, page 117] is overlooked in one of Smoryński's papers [51, page 394]. In the next section, we will correct Smoryński's defective proof (Simpson's proof is also defective, but he gives a glimpse of a "repair" at the very end of his paper, page 117).

By theorem 8.2, the ordinals less than Γ_0 can be defined recursively as follows.

Lemma 8.4 For every ordinal $\gamma < \Gamma_0$, either

(1) $\gamma = 0$, or

(2) $\gamma = \beta + \alpha$, for some ordinals $\alpha, \beta < \gamma$ such that $\alpha \leq \beta$, or

(3) $\gamma = \varphi(\alpha, \beta)$, for some ordinals $\alpha, \beta < \gamma$.

Proof. The proof follows immediately from theorem 8.2 by induction on n in the decomposition $\gamma = \varphi(\alpha_1, \beta_1) + \cdots + \varphi(\alpha_n, \beta_n)$. □

In case (3), we cannot guarantee that $\alpha \leq \beta$, and we have to consider the three subcases $\alpha < \beta$, $\alpha = \beta$, and $\alpha > \beta$. Actually, we can reduce these three cases to two if we replace $<$ by $\leq$.

This recursive representation of the ordinals $< \Gamma_0$ is the essence of the connection between Γ_0 and Kruskal's theorem explored in section 9.

Lemma 8.4 shows that every ordinal $\alpha < \Gamma_0$ *can be* represented in terms of 0, $+$, and φ, but this representation has some undesirable properties, namely that different notations can represent the same ordinal. In particular, for some $\alpha \leq \beta < \Gamma_0$, we may have $\beta = \varphi(\alpha, \beta)$ (which happens when $\beta \in Cr(\alpha')$). For example, $\epsilon_0 = \varphi(0, \epsilon_0)$ (since $\epsilon_0 = \varphi(1, 0)$). It would be desirable to have a representation similar to that given by lemma 8.2, but for a function ψ such that $\alpha < \psi(\alpha, \beta)$ and $\beta < \psi(\alpha, \beta)$, for all $\alpha, \beta < \Gamma_0$. Such a representation is possible, as shown in Schütte [46, Section 13.7, page 84-92]. The key point is to consider ordinals γ that are *maximal α-critical*, that is, maximal with respect to the property of belonging to some $Cr(\alpha)$.

Definition 8.5 An ordinal $\gamma \in \mathcal{O}$ is *maximal α-critical* iff $\gamma \in Cr(\alpha)$ and $\gamma \notin Cr(\beta)$ for all $\beta > \alpha$.

By proposition 6.22 and proposition 6.23, $\gamma \in Cr(\alpha)$ iff $\varphi_\beta(\gamma) = \gamma$ for all $\beta < \alpha$. Thus, γ is maximal α-critical iff $\varphi_\alpha(\gamma) \neq \gamma$. However, because φ_α is the ordering function of $Cr(\alpha)$, we know from proposition 6.9 that $\delta \leq \varphi_\alpha(\delta)$ for all δ, and so, γ is maximal α-critical iff $\gamma = \varphi_\alpha(\beta)$ for some $\beta < \gamma$. It follows from proposition 6.32 that for every principal additive number γ, there is some $\alpha \leq \gamma$ such that γ is maximal α-critical.

Definition 8.6 The function ψ_α is defined as the ordering function of the maximal α-critical ordinals.

We also define $\psi(\alpha, \beta)$ by letting $\psi(\alpha, \beta) = \psi_\alpha(\beta)$. It is possible to give a definition of ψ in terms of φ, as shown in Schütte [46].

Lemma 8.7 The function ψ defined such that

$$\psi(\alpha,\beta) = \begin{cases} \varphi(\alpha,\beta+1), & \text{if } \beta = \beta_0 + n \text{ and } \varphi(\alpha,\beta_0) = \beta_0, \\ & \text{for some } \beta_0 \text{ and } n \in \mathbf{N}; \\ \varphi(\alpha,\beta), & \text{otherwise.} \end{cases}$$

is the ordering function of the maximal α-critical ordinals for every α.

We list the following properties of ψ without proof, referring the reader to Schütte [46] for details.

Lemma 8.8 For every additive principal number γ, there are unique $\alpha, \beta \leq \gamma$ such that $\gamma = \psi(\alpha,\beta)$.

Lemma 8.9 (1) If $\gamma = \psi(\alpha,\beta)$, then $\alpha < \gamma$ iff γ is not strongly critical.

(2) $\beta < \psi(\alpha,\beta)$ for all α, β.

Lemma 8.10 $\psi(\alpha_1,\beta_1) < \psi(\alpha_2,\beta_2)$ holds iff either

(1) $\alpha_1 < \alpha_2$ and $\beta_1 < \psi(\alpha_2,\beta_2)$, or

(2) $\alpha_1 = \alpha_2$ and $\beta_1 < \beta_2$, or

(3) $\alpha_2 < \alpha_1$ and $\psi(\alpha_1,\beta_1) < \beta_2$.

It should be noted that the set of maximal α-critical ordinals is unbounded, but it is not closed, because the function ψ_α is *not* continuous. However, this is not a problem for representing the ordinals less than Γ_0.

Since Γ_0 is the least strongly critical ordinal, by lemma 8.9, we have the following corollary.

Lemm 8.11 For all $\alpha, \beta < \Gamma_0$, we have

(1) $\alpha < \psi(\alpha,\beta)$, and

(2) $\beta < \psi(\alpha,\beta)$.

Using lemma 8.8, we can prove another version of the normal form theorem 8.2 for the ordinal less than Γ_0, using ψ instead of φ.

Theorem 8.12 For every ordinal α such that $0 < \alpha < \Gamma_0$, there exist unique ordinals $\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_n$, $n \geq 1$, with $\alpha_i, \beta_i < \psi(\alpha_i,\beta_i) \leq \alpha$, $1 \leq i \leq n$, such that

(1) $\alpha = \psi(\alpha_1,\beta_1) + \cdots + \psi(\alpha_n,\beta_n)$, and

(2) $\psi(\alpha_1, \beta_1) \geq \ldots \geq \psi(\alpha_n, \beta_n)$.

The advantage of the reprentation given by theorem 8.12 is that it is now possible to design a system of notations where distinct notations represent distinct ordinals, and ψ satisfies the subterm property of lemma 8.11. Such a notation system will be given in section 11.

9 Kruskal's Theorem and Γ_0

The connection between Γ_0 and Kruskal's theorem lies in the fact that there is a close relationship between the embedding relation $\preceq$ on trees (definition 4.11) and the well-ordering $\leq$ on $\mathcal{O}(\Gamma_0)$ (recall that $\mathcal{O}(\Gamma_0)$ is the set of all ordinals $< \Gamma_0$).

We shall restrict our attention to tree domains, or equivalently assume that the set of labels contains a single symbol. Let T denote the set of all finite tree domains, which, for brevity are also called trees. In this case, by a previous remark, it is easy to show that $\preceq$ is a partial order. We shall exhibit a function $h : T \to \mathcal{O}(\Gamma_0)$ from the set of finite trees to the set of ordinals less that Γ_0, and show that h is (1). surjective, and (2). preserves order, that is, if $s \preceq t$, then $h(s) \leq h(t)$ (where $\preceq$ is the embedding relation defined in definition 4.11). It will follow that Kruskal's theorem (theorem 4.12) implies that $\mathcal{O}(\Gamma_0)$ is well-ordered by $\leq$, or put slightly differently, Kruskal's theorem implies the validity of transfinite induction on Γ_0. In turn, the provability of transfinite induction on large ordinals is known to be proof-theoretically significant. As first shown by Gentzen, one can prove the consistency of logical theories using transfinite induction on large ordinals. As a consequence, Kruskal's theorem in not provable in fairly strong logical theories, in particular some second-order theories for which transfinite induction up to Γ_0 is not provable.

We now give the definition of the function h mentioned above. In view of the recursive characterization of the ordinals $< \Gamma_0$, it is relatively simple to define a surjective function from T to $\mathcal{O}(\Gamma_0)$. However, making h order-preserving is more tricky. As a matter of fact, this is why lemma 8.3 is needed, but beware! Simpson defines a function h using five recursive cases, but points out at the end of his paper that there is a problem, due to the failure of the inequality $\beta + \alpha \leq \varphi(\alpha, \beta)$ [47, page 117]. Actually, a definition with fewer cases can be given, and Smoryński defines a function h using four recursive cases [51]. Unfortunately, Smoryński's definition also makes use of the erroneous inequality [51, page 394]. We give what we believe to be a repaired version of Smoryński's definition of h (using five recursive cases).

Remark. We do not know whether a definition using the function ψ of the previous section can be given. Certainly a surjective function can be defined using ψ, but the difficult

part is to insure monotonicity.

Definition 9.1 The function $h : T \to \mathcal{O}(\Gamma_0)$ from the set of finite trees to the set of ordinals less that Γ_0 is defined recursively as follows:

(0) $h(t) = 0$, when t is the one-node tree.

(1) $h(t) = h(t/1)$, if $rank(t) = 1$, i.e, the root of t has only one successor.

(2) $h(t) = \beta + \alpha$, if $rank(t) = 2$, where α is the least element of $\{h(t/1), h(t/2)\}$ and β is the largest.

(3) $h(t) = \varphi(\alpha, \beta)$, if $rank(t) = 3$, where $\alpha \leq \beta$ are the two largest elements of the set $\{h(t/1), h(t/2), h(t/3)\}$, and $\beta < \varphi(\alpha, \beta)$.

(4) $h(t) = \beta + \alpha$, if $rank(t) = 3$, where $\alpha \leq \beta$ are the two largest elements of the set $\{h(t/1), h(t/2), h(t/3)\}$, and $\beta = \varphi(\alpha, \beta)$.

(5) $h(t) = \varphi(\beta, \alpha)$, if $rank(t) \geq 4$, where $\alpha \leq \beta$ are the two largest elements of the set $\{h(t/1), h(t/2), \ldots, h(t/k)\}$, with $k = rank(t)$.

The following important theorem holds.

Theorem 9.2 The function $h : T \to \mathcal{O}(\Gamma_0)$ is surjective and monotonic, that is, for every two finite tree s, t, if $s \preceq t$, then $h(s) \leq h(t)$.

Proof (sketch). The fact that h is surjective follows directly from the recursive definition shown in lemma 8.4. Note that clause (1) and (4) are not needed for showing that h is a surjection, but they are needed to ensure that h is well defined and preserves order. By clause (0), $h(t) = 0$, for the one-node tree t. Clause (2) is used when $\gamma = \beta + \alpha$, with $\alpha, \beta < \gamma$ and $\alpha \leq \beta$. Clause (3) is used when $\gamma = \varphi(\alpha, \beta)$ with $\alpha, \beta < \gamma$ and $\alpha \leq \beta$, and clause (5) is used when $\gamma = \varphi(\beta, \alpha)$ with $\alpha, \beta < \gamma$ and $\alpha \leq \beta$.

The proof that if $s \preceq t$, then $h(s) \leq h(t)$ proceeds by cases, using induction on trees, corollary 6.30, and lemma 8.3. The only delicate case arises when $rank(s) = 2$, $rank(t) = 3$, and, assuming that $h(t/1) \geq h(t/2) \geq h(t/3)$ and $h(s/1) \geq h(s/2)$, we have $h(t/1) = \varphi(h(t/2), h(t/1))$, $s/1 \preceq t/1$ and $s/2 \preceq t/2$. By the induction hypothesis, $h(s/1) \leq h(t/1)$ and $h(s/2) \leq h(t/2)$, and since $h(s) = h(s/1) + h(s/2)$ and $h(t) = h(t/1) + h(t/2)$, we have $h(s) \leq h(t)$. If $h(t/1) < \varphi(h(t/2), h(t/1))$, then $h(t) = \varphi(h(t/2), h(t/1))$, and by proposition 8.3, $h(s) = h(s/1) + h(s/2) \leq h(t/1) + h(t/2) \leq \varphi(h(t/2), h(t/1)) = h(t)$. The other cases are left to the reader. □

Theorem 9.2 implies that there exist total orderings of order type Γ_0 extending the partial order $\preceq$ on (finite) trees. DeJongh and Parikh [6] proved that the maximum (sup)

of all the total extensions is attained, and they computed the maximum for certain of the (Higman) orderings. The ordinals associated with various orderings on trees arising in the theory of rewriting systems have been investigated by Dershowitz and Okada [9], Okada and Takeuti [38], and Okada [37, 39, 40].

Theorem 9.2 also has the following important corollary.

Lemma 9.3 Kruskal's theorem implies that $\mathcal{O}(\Gamma_0)$ is well-ordered by $\leq$.

Proof. Assume that there is some infinite sequence $(\alpha_i)_{i\geq 1}$ of ordinals in $\mathcal{O}(\Gamma_0)$ such that $\alpha_{i+1} < \alpha_i$ for all $i \geq 1$. By theorem 9.2, since h is surjective, there is an infinite sequence of trees $(t_i)_{i\geq 1}$ such that $h(t_i) = \alpha_i$ for all $i \geq 1$. By Kruskal's theorem (theorem 4.12), there exist $i, j > 0$ such that $i < j$ and $t_i \preceq t_j$. By theorem 9.2, we have $\alpha_i = h(t_i) \leq h(t_j) = \alpha_j$, contradicting the fact that $\alpha_j < \alpha_i$. Hence, $\mathcal{O}(\Gamma_0)$ is well-ordered by $\leq$. □

Let us denote by $WO(\Gamma_0)$ the property that $\mathcal{O}(\Gamma_0)$ is well-ordered by $\leq$, and by $WQO(T)$ the property that the embedding relation $\preceq$ is a *wqo* on the set T of finite trees. $WQO(T)$ is a formal statement of Kruskal's theorem.

For every formal system $\mathcal{S}$, if the proof that $(WQO(T) \supset WO(\Gamma_0))$ (given in lemma 9.3) can be formalized in $\mathcal{S}$ and $WO(\Gamma_0)$ is not provable in $\mathcal{S}$, then $WQO(T)$ is not provable in $\mathcal{S}$. In the next section, we briefly describe some subsystems of 2^{nd}-order arithmetic in which Kruskal's theorem and its miniature versions are not provable.

10 The Subsystems ACA_0, ATR_0, Π^1_1-CA_0, of Second-Order Arithmetic

Harvey Friedman has shown that $WO(\Gamma_0)$ is not provable in some relatively strong subsystems of 2^{nd}-order arithmetic, and therefore, Kruskal's theorem is not provable in such systems. Friedman also proved similar results for some finite (first-order) miniaturizations of Kruskal's theorem. In particular, these first-order versions of Kruskal's theorem are not provable in Peano's arithmetic, since transfinite induction up to ϵ_0 is not provable in Peano's arithmetic, due to a result of Gentzen. We now provide some details on these subsystems of 2^{nd}-order arithmetic.

Second-order arithmetic can be formulated over a two-sorted language with *number variables* $(m, n, \ldots)$ and *set variables* $(X, Y, \ldots)$. We define *numerical terms* as terms built up from number variables, the constant symbols 0, 1, and the function symbols + (addition) and · (multiplication). An *atomic formula* is either of the form $t_1 \doteq t_2$, or $t_1 < t_2$, or $t_1 \in X$, where t_1 and t_2 are numerical terms. A *formula* is built up from atomic formulae using

$\wedge, \vee, \supset, \equiv, \neg$, number quantifiers $\forall n, \exists n$, and set quantifiers $\forall X, \exists X$. We say that a formula is *arithmetical* iff it does not contain set quantifiers.

All systems of second-order arithmetic under consideration include standard axioms stating that $\langle \mathbf{N}, 0, 1, +, \cdot, < \rangle$ is an ordered semi-ring. The real power of a system of second-order arithmetic is given by the form of its *induction axioms*, and the form of its *comprehension axioms*.

For the systems under consideration, the induction axiom is

$$[0 \in X \wedge \forall m(m \in X \supset m + 1 \in X)] \supset \forall n(n \in X),$$

where X is a set variable. This form of induction is often called *restricted induction*, in contrast with the principle of *full induction* stated as

$$[\varphi(0) \wedge \forall m(\varphi(m) \supset \varphi(m + 1))] \supset \forall n \varphi(n),$$

where φ is an arbitrary 2^{nd}-order formula. Apparently, Friedman initiated the study of subsystems of 2^{nd}-order arithmetic with restricted induction (this explains the subscript 0 after the name of the systems ACA, ATR, or Π^1_1-CA).

The system Π^1_∞-CA_0, also known as Z_2, or *second-order arithmetic*, has comprehension axioms of the form

$$\exists X \forall n(n \in X \equiv \varphi(n)),$$

where φ is any 2^{nd}-order formula φ in which X is not free. This is a very powerful form of comprehension axioms. Susbystems of Z_2 are obtained by restricting the class of formulae for which comprehension axioms hold.

The system ACA_0 is obtained by restricting the comprehension axioms to *arithmetical formulae* in which X is not free (ACA stands for Arithmetical Comprehension Axioms). It turns out that ACA_0 is a conservative extension of (first-order) Peano Arithmetic (PA). A weak form of König's lemma is provable in ACA_0, and a fairly smooth theory of continuous functions and of sequential convergence can be developed. For example, Friedman proved that the Bolzano/Weierstrass theorem (every bounded sequence of real numbers contains a convergent subsequence) is provable in ACA_0. In fact, Friedman proved the stronger result that no set existence axioms weaker than those of ACA_0 are sufficient to establish the Bolzano/Weierstrass theorem. For details, the reader is referred to Simpson [48].

The system ATR_0 contains axioms stating that arithmetical comprehension can be iterated along any countable well ordering (ATR stands for Arithmetical Transfinite Recursion). A precise formulation of the axiom ATR can be found in Friedman, McAloon, and

Simpson [16] (see also Feferman [14]), but it is not essential here. The system ATR_0 permits a convenient development of a large part of ordinary mathematics, including, the theory of continuous functions, the Riemann integral, the theory of countable fields, the topology of complete separable metric spaces, the structure theory of separable Banach spaces, a good theory of countable well orderings, Borel sets, analytic sets, and more.

The system $\Pi^1_1\text{-}CA_0$ is obtained by allowing comprehension axioms in which φ is any Π^1_1-formula in which X is not free. This is a system even stronger that ATR_0, whose axioms imply many mathematical results in the realm of algebra, analysis, classical descriptive set theory, and countable combinatorics.

The systems ACA, ATR and $\Pi^1_1\text{-}CA$ allow full induction rather than restricted induction. It might be interesting to mention that the least ordinals for which transfinite induction cannot be proved in ACA_0 and ATR_0 are respectively ϵ_0 and Γ_0. Such an ordinal has also be determined for $\Pi^1_1\text{-}CA_0$, but the notation system required to describe it is beyond the scope of this paper. In contrast, the least ordinals for which transfinite induction cannot be proved in ACA and ATR are respectively ϵ_{ϵ_0} and Γ_{ϵ_0}.

We now return to the connections with Γ_0 and Kruskal's theorem. Friedman has shown that $WO(\Gamma_0)$ is not provable in ATR_0 (Friedman, McAloon, and Simpson [16]). He also showed that $(WQO(T) \supset WO(\Gamma_0))$ is provable in ACA_0. Since ACA_0 is a subsystem of ATR_0, we conclude that $WQO(T)$ is not provable in ATR_0. This is already quite remarkable, considering that a large part of ordinary mathematics can be done in ATR_0. But Friedman also proved that the miniature version $LWQO(T)$ of Kruskal theorem given in theorem 5.1 is not provable in ATR_0, an even more remarkable result. The proof of this last result is given in Simpson [47].

There is one more "tour de force" of Friedman that we have not mentioned! Harvey Friedman has formulated an extension of the miniature version of Kruskal's theorem (using a gap condition), and proved that this version of Kruskal's theorem is not provable in $\Pi^1_1\text{-}CA_0$. The proof can be found in Simpson [47]. There are also some connections bewteen this last version of Kruskal's theorem and certain ordinal notations due to Takeuti known as ordinals diagrams. These connections ae investigated in Okada and Takeuti [38], and Okada [39, 40].

11 A Brief Introduction to Term Orderings

This section is a brief introduction to term orderings. These orderings play an important role in computer science, because they are the main tool for showing that sets of rewrite rules are finite terminating (Noetherian). In turn, Noetherian sets of rewrite rules play a

fundamental role in automated deduction in equational logic. Indeed, one of the major techniques in equational logic is to complete a given set of equations E to produce an equivalent set R of rewrite rules which has some "good" properties, namely to be confluent and Noetherian. A number of procedures that attempt to produce such a set R of rewrite rules from a set E of equations have been designed. The first such procedure is due to Knuth and Bendix [27], but there are now many kinds of completion procedures. For more details on completion procedures, we refer the reader to Dershowitz [11] and Bachmair [2].

There are many classes of term orderings, but an important class relevant to our considerations is the class of simplification orderings, because Kruskal's theorem can be used to prove the well-foundedness of these orderings. For a comprehensive study of term orderings, the reader is referred Dershowitz's excellent survey [7] and to Dershowitz's fundamental paper [8].

Given a set of labels Σ, the notion of a tree was defined in definition 4.2. When considering rewrite rules, we usually assume that Σ is a ranked alphabet, that is, that there is a ranking function $r : \Sigma \to \mathbf{N}$ assigning a natural number $r(f)$, the *rank* (or *arity*) of f, to every $f \in \Sigma$. We also have a countably infinite set $\mathcal{X}$ of variables, with $r(x) = 0$ for every $x \in \mathcal{X}$, and we let $T_\Sigma(\mathcal{X})$ be the set of all trees (also called Σ-*terms*, or *terms*) $t \in T_{\Sigma \cup \mathcal{X}}$ such that, for every tree address $u \in dom(t)$, $r(t(u)) = rank(t/u)$. In other words, the rank of the label of u is equal to the rank of t/u (see definition 4.3), the number of immediate successors of u.

Given a tree t, we let $Var(t) = \{x \in \mathcal{X} \mid \exists u \in dom(t),\ t(u) = x\}$ denote the set of variables occurring in t. A *ground term* t is a term such that $Var(t) = \emptyset$.

Definition 11.1 A *set of rewrite rules* is a binary relation $R \subseteq T_\Sigma(\mathcal{X}) \times T_\Sigma(\mathcal{X})$ such that $Var(r) \subseteq Var(l)$ whenever $\langle l, r \rangle \in R$.

A rewrite rule $\langle l, r \rangle \in R$ is usually denoted as $l \to r$. The notions of tree replacement and substitution are needed for the definition of the rewrite relation induced by a set of rewrite rules.

Definition 11.2 Given two trees t_1 and t_2 and a tree address u in t_1, the *result of replacing* t_2 at u in t_1, denoted by $t_1[u \leftarrow t_2]$, is the function whose graph is the set of pairs

$$\{(v, t_1(v)) \mid v \in dom(t_1),\ u \text{ is not a prefix of } v\} \cup \{(uv, t_2(v)) \mid v \in dom(t_2)\}.$$

Definition 11.3 A *substitution* is a function $\sigma : \mathcal{X} \to T_\Sigma(\mathcal{X})$, such that, $\sigma(x) \neq x$ for only finitely many $x \in \mathcal{X}$. Since $T_\Sigma(\mathcal{X})$ is the free Σ-algebra generated by $\mathcal{X}$, every substitution $\sigma : \mathcal{X} \to T_\Sigma(\mathcal{X})$ has a unique homomorphic extension $\widehat{\sigma} : T_\Sigma(\mathcal{X}) \to T_\Sigma(\mathcal{X})$. In the sequel, we will identify σ and its homomorphic extension $\widehat{\sigma}$, and denote $\widehat{\sigma}(t)$ as $t[\sigma]$.

Definition 11.4 Given a substitution σ, the *domain* of σ is the set of variables $D(\sigma) = \{x \mid \sigma(x) \neq x\}$. Given a substitution σ, if its domain is the set $\{x_1, \ldots, x_n\}$, and if $t_i = \sigma(x_i)$, $1 \leq i \leq n$, then σ is also denoted by $[t_1/x_1, \ldots, t_n/x_n]$.

Definition 11.5 A substitution σ is a *renaming* iff $\sigma(x)$ is a variable for every $x \in D(\sigma)$, and σ is injective. Let $R \subseteq T_\Sigma(\mathcal{X}) \times T_\Sigma(\mathcal{X})$ be a set of rewrite rules. A rewrite rule $s \to t$ is a *variant* of a rewrite rule $u \to v \in R$ iff there is some renaming ρ with domain $Var(u) \cup Var(v)$ such that $s = u[\rho]$ and $t = v[\rho]$.

Definition 11.6 Let $\longrightarrow$ be a binary relation $\longrightarrow \subseteq T_\Sigma(\mathcal{X}) \times T_\Sigma(\mathcal{X})$. (i) The relation $\longrightarrow$ is *monotonic* (or *stable under the algebra structure*) iff for every two terms s, t and every function symbol $f \in \Sigma$, if $s \longrightarrow t$ then $f(\ldots, s, \ldots) \longrightarrow f(\ldots, t, \ldots)$.

(ii) The relation $\longrightarrow$ is *stable* (under substitution) if $s \longrightarrow t$ implies $s[\sigma] \longrightarrow t[\sigma]$ for every substitution σ.

Definition 11.7 Let $R \subseteq T_\Sigma(\mathcal{X}) \times T_\Sigma(\mathcal{X})$ be a set of rewrite rules. The relation $\longrightarrow_R$ over $T_\Sigma(\mathcal{X})$ is defined as the smallest stable and monotonic relation that contains R. This is the *rewrite relation* associated with R.

This relation is defined explicitly as follows: Given any two terms $t_1, t_2 \in T_\Sigma(\mathcal{X})$, then

$$t_1 \longrightarrow_R t_2$$

iff there is some variant $l \to r$ of some rule in R, some tree address α in t_1, and some substitution σ, such that

$$t_1/\alpha = l[\sigma], \quad \text{and} \quad t_2 = t_1[\alpha \leftarrow r[\sigma]].$$

We say that a rewrite system R is Noetherian iff the relation $\longrightarrow_R$ associated with R is Noetherian.

Now, our goal is to describe some orderings that will allow us to prove that sets of rewrite rules are Noetherian. First, it is convenient to introduce the concept of a strict ordering.

Definition 11.8 A *strict ordering* (or *strict order*) $\prec$ on a set A is a transitive and irreflexive relation (for all a, $a \not\prec a$.)

Given a preorder (or partial order) $\preceq$ on a set A, the strict ordering $\prec$ associated with $\preceq$ is defined such that $s \prec t$ iff $s \preceq t$ and $t \not\preceq s$. Conversely, given a strict ordering $\prec$,

the partial ordering $\preceq$ associated with $\prec$ is defined such that $s \preceq t$ iff $s \prec t$ or $s = t$. The converse of a strict ordering $\prec$ is denoted as $\succ$.

We now introduce the important concepts of simplification ordering, and reduction ordering. Let Σ be a set of labels (in most cases, a ranked alphabet).

Definition 11.9 A strict order $\prec$ on T_Σ satisfying conditions

(1) $s \prec f(\ldots, s, \ldots)$, and

(2) $f(\ldots) \prec f(\ldots, s, \ldots)$,

is said to have the *subterm property* and the *deletion property*.

A *simplification ordering* $\prec$ is a strict ordering that is monotonic and has the subterm and deletion property.[1]

A *reduction ordering* $\prec$ is a strict ordering that is monotonic, stable under substitution, and such that $\succ$ is well-founded.

With a slight abuse of language, we will also say that the converse $\succ$ of a strict ordering $\prec$ is a simplification ordering (or a reduction ordering). The importance of term orderings is shown by the next fundamental result.

Lemma 11.10 A set of rules R is Noetherian if and only if there exists a reduction ordering $\succ$ on $T_\Sigma(\mathcal{X})$ such that $l \succ r$ for every $l \to r \in R$.

Unfortunately, it is undecidable in general if an arbitrary system R is Noetherian since it is possible to encode Turing machines using a system of two rewrite rules, and this would imply the decidability of the halting problem (see Dershowitz [7]). The importance of simplification orderings is shown by the next theorem.

Theorem 11.11 (Dershowitz) If Σ is finite, then every simplification ordering on T_Σ is well-founded.

Proof. This is a consequence of proposition 4.8, which uses Kruskal's tree theorem. □

In practice, we want theorem 11.11 to apply to simplification orderings on $T_\Sigma(\mathcal{X})$, but since $\mathcal{X}$ is infinite, there is a problem. However, we are saved because we usually only care about terms arising in derivations.

---

[1] When Σ is a ranked alphabet, the deletion property is superfluous.

Definition 11.12 An ordering $\succ$ is *well-founded for derivations* iff $\succ \cap \xrightarrow{*}_R$ is well-founded for every finite rewrite system R.

Since $Var(r) \subseteq Var(l)$ for every $l \to r \in R$, every derivation of a finite rewrite system involves only finitely many symbols. Thus, as corollary of the above theorem we have:

Corollary 11.13 (Dershowitz) Every simplification ordering is well-founded for derivations.

Warning: There exists rewrite systems whose termination cannot be shown by any total simplification ordering as shown by the following example.

Example 11.14

$$f(a) \to f(b)$$
$$g(b) \to g(a)$$

Next, we are going to describe two important classes of simplification orderings, the recursive path ordering, and the lexicographic path ordering. But first, we need to review the definitions of the lexicographic ordering and the multiset ordering.

Definition 11.15 Given n partially ordered sets $(S_i, \prec_i)$ (where each $\prec_i$ is a strict order, $n > 1$), the *lexicographic order* $\prec_{lex}$ on the set $S_1 \times \cdots \times S_n$ is defined as follows. Let $\langle a_1, \ldots, a_n \rangle$ and $\langle b_1, \ldots, b_n \rangle$ be members of $S_1 \times \cdots \times S_n$. Then

$$\langle a_1, \ldots, a_n \rangle \prec_{lex} \langle b_1, \ldots, b_n \rangle$$

if and only if there exists some i, $1 \leq i \leq n$, such that $a_i \prec_i b_i$, and $a_j = b_j$ for all j, $1 \leq j < i$.

We now turn to multiset orderings. Multiset orderings have been investigated by Dershowitz and Manna [10], and Jouannaud and Lescanne [24].

Definition 11.16 Given a set A, a *multiset* over A is an unordered collection of elements of A which may have multiple occurrences of identical elements. More formally, a multiset over A is a function $M : A \to \mathbf{N}$ (where $\mathbf{N}$ is the set of natural numbers) such that an element $a \in A$ has exactly n occurrences in M iff $M(a) = n$. In particular, a does not belong to M when $M(a) = 0$, and we say that $a \in M$ iff $M(a) > 0$.

The *union* of two multisets M_1 and M_2, denoted by $M_1 \cup M_2$, is defined as the multiset M such that for all $a \in A$, $M(a) = M_1(a) + M_2(a)$.

Let $(S, \prec)$ be a partially ordered set (where $\prec$ is a strict order), let M be some finite multiset of objects from S, and finally let $n, n'_1, \ldots, n'_k \in S$. Define the relation $\Leftarrow_m$ on finite multisets as

$$M \cup \{n'_1, \ldots, n'_k\} \Leftarrow_m M \cup \{n\},$$

where $k \geq 0$ and $n'_i \prec n$ for all i, $1 \leq i \leq k$.

The multiset ordering $\prec_{M(S)}$ is simply the transitive closure $\overset{+}{\Leftarrow}_m$.

In other words, $N' \prec_{M(S)} N$ iff N' is produced from a finite multiset N by removing one or more elements and replacing them with any finite number of elements, each of which is strictly smaller than at least one element removed. For example, $\{4,4,3,3,1\} \prec \{5,3,1,1\}$, where $\prec$ is the multiset ordering induced by the ordering $<$ of the natural numbers.

It is easy to show that for any partially ordered set $(S, \preceq)$, we have associated partially ordered sets $(M(S), \preceq_{M(S)})$ (where $M(S)$ is the set of all finite multisets of members of S), and $(S^n, \preceq_{lex})$ for $n > 0$. Furthermore $\preceq$ is total (respectively, well-founded) iff $\preceq_{lex}$ (for any n) is total (respectively, well-founded).

Using König's lemma, we can also show the following useful result.

Lemma 11.17 If $\preceq$ is well-founded (respectively, total) on S, then $\preceq_{M(S)}$ is well-founded (respectively, total) on $M(S)$.

There is an interesting connection between the multiset ordering and ordinal exponentiation. Given a well ordering $\preceq$ on a set S, it is well know that there is a unique ordinal α and a unique order-preserving bijection $\varphi : S \to \alpha$.

The connection is that $(M(S), \prec_{M(S)})$ is order-isomorphic to ω^α. Indeed, the function $\psi: M(S) \to \omega^\alpha$ defined such that $\psi(\emptyset) = 0$, and

$$\psi(\{m_1, \ldots, m_k\}) = \omega^{\varphi(m_1)} + \cdots + \omega^{\varphi(m_k)},$$

where $\varphi(m_1) \geq \ldots \geq \varphi(m_k)$ is the nonincreasing sequence enumerating $\varphi(\{m_1, \ldots, m_k\})$,[2] is easily shown to be an order-isomorphism.

The lexicographic ordering and the multiset ordering can also be defined for preorders. This generalization will be needed for defining *rpo* and *lpo* orderings based on preorders.

Definition 11.18 Given n preordered sets $(S_i, \preceq_i)$ $(n > 1)$, the *lexicographic preorder* $\preceq_{lex}$ on the set $S_1 \times \cdots \times S_n$ is defined as follows:

$$\langle a_1, \ldots, a_n \rangle \preceq_{lex} \langle b_1, \ldots, b_n \rangle$$

[2] In the theory of ordinals, the sum $\omega^{\varphi(m_1)} + \cdots + \omega^{\varphi(m_k)}$ is a *natural sum*.

if and only if there exists some i, $1 \leq i \leq n$, such that $a_i \preceq_i b_i$, and $a_j \approx_j b_j$ for all j, $1 \leq j < i$.[3]

Definition 11.19 Let $(S, \preceq)$ be a preordered set, let M be some finite multiset of objects from S, and finally let $n, n'_1, \ldots, n'_k \in S$. Define the relation $\Leftarrow_m$ on finite multisets as

$$M \cup \{n'_1, \ldots, n'_k\} \Leftarrow_m M \cup \{n\},$$

where either $k = 1$ and $n \approx n'_1$, or $k \geq 0$ and $n'_i \prec n$ for all i, $1 \leq i \leq k$.[4]

The multiset preorder $\preceq_{M(S)}$ is the transitive closure $\overset{+}{\Leftarrow}_m$.

Two finite multisets M_1 and M_2 are equivalent ($M_1 \approx_{M(S)} M_2$) iff they have the same number of elements, and every element of M_1 is equivalent to some element of M_2 and vice versa. It is easy to show that for any preordered set $(S, \preceq)$ we have associated preordered sets $(M(S), \preceq_{M(S)})$ (where $M(S)$ is the set of all finite multisets of members of S), and $(S^n, \preceq_{lex})$ for $n > 0$. Furthermore $\preceq$ is total (respectively, well-founded) iff $\preceq_{lex}$ (for any n) is total (respectively, well-founded).

Using König's lemma, we can also show that lemma 11.17 holds for preorders.

Lemma 11.20 If $\preceq$ is a well-founded preorder (respectively, total) on S, then $\preceq_{M(S)}$ is well-founded (respectively, total) on $M(S)$.

A naive ordering on terms based on the notion of lexicographic order is as follows.

For any given ordering $\succ$ on Σ we say that

$$s = f(s_1, \ldots, s_n) \succ^{tlex} g(t_1, \ldots, t_m) = t$$

iff either

(i) $f \succ g$; or

(ii) $f = g$ and $\langle s_1, \ldots, s_n \rangle \succ^{tlex}_{lex} \langle t_1, \ldots, t_n \rangle$,

where $\succ^{tlex}_{lex}$ is the lexicographic extension of $\succ^{tlex}$ to n-tuples of terms (the success of this recursive definition depends on the fact that we use the lexicographic extension over terms *smaller* than s and t).

It is easy to show by structural induction on terms that *tlex* is total on ground terms whenever the $\succ$ is total on Σ, but it has a severe defect: it is not well-founded. For example,

---

[3] As usual, the equivalence $\approx$ associated with a preorder $\preceq$ is defined such that $a \approx b$ iff $a \preceq b$ and $b \preceq a$.

[4] As usual, given a preorder $\preceq$, the strict order $\prec$ is defined such that $a \prec b$ iff $a \preceq b$ and $b \not\preceq a$.

if $a \succ f$ then we have $a \succ^{tlex} fa \succ^{tlex} f^2a \succ^{tlex} \ldots$. The problem arises since it is possible for a term to be strictly smaller than one of its subterms.

The most powerful forms of reduction orderings are based on the relative syntactic simplicity of two terms, i.e., on the notion of a simplification ordering. Although there are many types of simplification orderings, one of the most elegant and useful is the *recursive path ordering*, for short, *rpo*.

Definition 11.21 Let $\preceq$ be a preorder on Σ. The *recursive path ordering* $\preceq_{rpo}$ on $T_\Sigma(\mathcal{X})$, for short, *rpo*, is defined below. Actually, we give a simultaneous recursive definition of $\succeq_{rpo}$, $\succ_{rpo}$, and $\approx_{rpo}$, where $s \succ_{rpo} t$ iff $s \succeq_{rpo} t$ and $s \not\preceq_{rpo} t$, and $s \approx_{rpo} t$ iff $s \succeq_{rpo} t$ and $s \preceq_{rpo} t$.

Then, $f(s_1, \ldots, s_n) \succeq_{rpo} g(t_1, \ldots, t_m)$ holds iff one of the conditions below holds:

(i) $f \approx g$ and $\{s_1, \ldots, s_n\} \succeq_{rpo}^{mult} \{t_1, \ldots, t_m\}$; or
(ii) $f \succ g$ and $f(s_1, \ldots, s_n) \succ_{rpo} t_i$ for *all* i, $1 \leq i \leq m$; or
(iii) $s_i \succeq_{rpo} g(t_1, \ldots, t_m)$ for *some* i, $1 \leq i \leq n$,

where $\succeq_{rpo}^{mult}$ is the extension of $\succeq_{rpo}$ to multisets,[5]

Note that since the preorder $\preceq$ is only defined on Σ, variables are regarded as incomparable symbols. In (ii), the purpose of the condition $f(s_1, \ldots, s_n) \succ_{rpo} t_i$ for *all* i, is to insure that $f(s_1, \ldots, s_n) \succ_{rpo} g(t_1, \ldots, t_m)$.

Theorem 11.22 (Dershowitz, Lescanne) The relation $\succ_{rpo}$ is a simplification ordering stable under substitution. Furthermore, if the strict order $\succ$ is well-founded on Σ, then $\succ_{rpo}$ is well-founded, even when Σ is infinite.

Proof sketch. Proving that *rpo* is a simplification ordering is laborious, especially transitivity. The complete proof can be found in Dershowitz [8]. In order to prove that $\succ_{rpo}$ is well-founded when $\succ$ is well-founded on Σ, it is tempting to apply proposition 4.8 to the preorders $\ll$ and $\preceq_{rpo}$, where $\ll$ is defined such that $s \ll t$ iff $root(s) \preceq root(t)$, since the conditions of this lemma hold. Unfortunately, $\preceq$ is not a *wqo*. However, we can use the idea from theorem 4.10 to extend $\preceq$ to a total well-founded ordering $\leq$. Then, by theorem 4.7, the embedding preorder $\preceq_\leq$ induced by $\leq$ (see definition 4.6) is a *wqo*, and thus, it is well-founded. We can now apply proposition 4.8, which shows that $\leq_{rpo}$ (the *rpo* induced by $\leq$) is well-founded. Finally, we prove by induction on terms that $\leq_{rpo}$ contains $\preceq_{rpo}$, which proves that $\succ_{rpo}$ itself is well-founded. □

[5] Other authors define $\succ_{rpo}^{mult}$ as the multiset extension of the strict order $\succ_{rpo}$, and $s \succeq_{rpo}^{mult} t$ iff $s \succ_{rpo}^{mult} t$ or $s = t$. Our definition is more general.

A proof not involving Kruskal's theorem, but using Zorn's lemma, is given in Lescanne [29]. Of course, a strict order on a finite set is always a *wqo*, and the significance of the second part of the theorem is that it holds even when Σ is infinite.

Example 11.23 Consider the following set of rewrite rules to convert a proposition to disjunctive normal form:

$$\begin{aligned}
\neg(P \vee Q) &\longrightarrow \neg P \wedge \neg Q,\\
\neg(P \wedge Q) &\longrightarrow \neg P \vee \neg Q,\\
P \wedge (Q \vee R) &\longrightarrow (P \wedge Q) \vee (P \wedge R),\\
(P \vee Q) \wedge R &\longrightarrow (P \wedge R) \vee (Q \wedge R),\\
\neg\neg P &\longrightarrow P,\\
P \vee P &\longrightarrow P,\\
P \wedge P &\longrightarrow P.
\end{aligned}$$

This system can be easily shown to be Noetherian using the *rpo* induced by the following ordering on the set of operators: $\neg \succ \wedge \succ \vee$.

It is possible to show that $\succeq_{rpo}$ is total on ground terms whenever $\succ$ is total on Σ. It is also possible to define reduction orderings which are total on ground terms; the problem with $\succeq_{rpo}$ is that it is not a partial order in general, but only a preorder, i.e., the equivalence relation $\approx_{rpo}$ is not necessarily the identity. For example, for any $\succ$ we have $f(a,b) \approx_{rpo} f(b,a)$ but clearly $f(a,b) \neq f(b,a)$. It is easy to show by structural induction on terms, and using only clause (i) of the definition of *rpo* that for any two ground terms $s = f(s_1, \ldots, s_n)$ and $t = g(t_1, \ldots, t_m)$, we have $s \approx_{rpo} t$ iff $f \approx g$ and $s_i \approx_{rpo} t_{\pi(i)}$, for $1 \leq i \leq n$, where π is some permutation of the set $\{1, \ldots, n\}$. (In other words, $s \approx_{rpo} t$ iff s and t are equal up to equivalence of symbols, and up to the permutation of the order of the terms under each function symbol, where the permutation of subterms arises by the comparison of multisets of subterms in clause (i) of the definition.)

This motivates the following definition.

Definition 11.24 For any ordering $\succ$ on Σ, let the term ordering $\succ_{rpol}$ be defined such that $s \succ_{rpol} t$ iff either $s \succ_{rpo} t$ or s and t are ground, $s \approx_{rpo} t$, and $s \succ^{tlex} t$.

Clearly for any total $\succ$ on Σ this is a reduction ordering total on ground terms, since $\succeq_{rpo}$ is total on ground terms and if $s \succeq_{rpo} t$ and $s \preceq_{rpo} t$ then, since $\succ^{tlex}$ is total on ground terms, we must have either $s \succ^{tlex} t$ or $s \prec^{tlex} t$.

Thus, any time the underlying ordering on Σ is total we have a total ordering on T_Σ, even though the ordering may not be total on $T_\Sigma(\mathcal{X})$. This is a major problem with

term orderings: in order to preserve stability under substitution, they must treat variables as incomparable symbols. Thus equations such as commutative axioms (e.g. $f(x,y) \doteq f(y,x)$) can never be oriented.

Warning: It is possible that for R and S rewrite systems on disjoint sets of function (and constant) symbols, both R and S are Noetherian, but $R \cup S$ is not, as shown by the following example due to Toyama.

Example 11.25

$$\begin{aligned} R &= \{f(0,1,z) \to f(z,z,z)\} \\ S &= \{g(x,y) \to x \\ &\quad\ g(x,y) \to y\} \end{aligned}$$

Observe that the term $f(g(0,1),g(0,1),g(0,1))$ rewrites to itself:

$$\begin{aligned} f(g(0,1),g(0,1),g(0,1)) &\longrightarrow f(0,g(0,1),g(0,1)) \\ &\longrightarrow f(0,1,g(0,1)) \\ &\longrightarrow f(g(0,1),g(0,1),g(0,1)). \end{aligned}$$

Another interesting kind of term ordering is the *lexicographic path ordering* due to Kamin and Levy.

Definition 11.26 Let $\preceq$ be a preorder on Σ. The *lexicographic path ordering* $\preceq_{lpo}$ on $T_\Sigma(\mathcal{X})$, for short, *lpo*, is defined below. Actually, we give a simultaneous recursive definition of $\succeq_{lpo}$, $\succ_{lpo}$, and $\approx_{lpo}$, where $s \succ_{lpo} t$ iff $s \succeq_{lpo} t$ and $s \not\preceq_{lpo} t$, and $s \approx_{lpo} t$ iff $s \succeq_{lpo} t$ and $s \preceq_{lpo} t$.

Then, $f(s_1, \ldots, s_n) \succeq_{lpo} g(t_1, \ldots, t_m)$ holds iff one of the conditions below holds:

(i) $f \approx g$, $s_1 \approx_{lpo} t_1, \ldots, s_{i-1} \approx_{lpo} t_{i-1}, s_i \succeq_{lpo} t_i$, and $s \succ_{lpo} t_{i+1}, \ldots, s \succ_{lpo} t_n$, for some i, $1 \leq i \leq n$, with $s = f(s_1, \ldots, s_n)$ and $m = n$; or

(ii) $f \succ g$ and $f(s_1, \ldots, s_n) \succ_{lpo} t_i$ for *all* i, $1 \leq i \leq m$; or

(iii) $s_i \succeq_{lpo} g(t_1, \ldots, t_m)$ for *some* i, $1 \leq i \leq n$.

Note that since the preorder $\preceq$ is only defined on Σ, variables are regarded as incomparable symbols. Also, condition (i) is sometimes stated as:

(i') $f \approx g$, $\langle s_1, \ldots, s_n \rangle \succeq^{lex}_{lpo} \langle t_1, \ldots, t_n \rangle$, $m = n$, and $f(s_1, \ldots, s_n) \succ_{lpo} t_i$ for all i, $1 \leq i \leq n$, where $\succeq^{lex}_{lpo}$ is the lexicographic extension of $\succeq_{lpo}$ on n-tuples.[6]

[6] Other authors define $\succ^{lex}_{lpo}$ as the lexicographic extension of the strict order $\succ_{lpo}$, and $s \succeq^{lex}_{lpo} t$ iff $s \succ^{lex}_{lpo} t$ or $s = t$. Our definition is more general.

It is easily seen that (i) and (i') are equivalent. In (i), the purpose of the conditions $s \succ_{lpo} t_{i+1}, \ldots, s \succ_{lpo} t_n$ is to insure that $f(s_1, \ldots, s_n) \succ_{lpo} g(t_1, \ldots, t_m)$ iff $s_i \succ_{lpo} t_i$. Similarly, in (ii), the purpose of the condition $f(s_1, \ldots, s_n) \succ_{lpo} t_i$ for *all* i, is to insure that $f(s_1, \ldots, s_n) \succ_{lpo} g(t_1, \ldots, t_m)$.

Theorem 11.27 (Kamin, Levy) The relation $\succ_{lpo}$ is a simplification ordering stable under substitution. Furthermore, if the strict order $\succ$ is well-founded on Σ, and equivalent symbols have the same rank, then $\succ_{lpo}$ is well-founded, even when Σ is infinite.

Proof. The proof uses the techniques used in theorem 11.22 (Kruskal's theorem). □

As in the previous theorem on *rpo*, the significance of the second part of the theorem is that it holds even when Σ is infinite.

Example 11.28 Consider the following set of rewrite rules for free groups (Knuth and Bendix [27]).

$$\begin{aligned}
(x * y) * z &\longrightarrow x * (y * z),\\
1 * x &\longrightarrow x,\\
I(x) * x &\longrightarrow 1,\\
I(x) * (x * y) &\longrightarrow y,\\
I(1) &\longrightarrow 1,\\
x * 1 &\longrightarrow x,\\
I(I(x)) &\longrightarrow x,\\
x * I(x) &\longrightarrow 1,\\
x * (I(x) * y) &\longrightarrow y,\\
I(x * y) &\longrightarrow I(y) * I(x).
\end{aligned}$$

This system can be easily shown to be Noetherian using the *lpo* induced by the following ordering on the set of operators: $I \succ * \succ 1$.

It is possible to combine *lpo* and *rpo* (Lescanne [32]). It is also possible to define *semantic path orderings* (Kamin, Levy), as opposed to the above *precedence orderings*. Semantic path orderings use orderings on T_Σ rather than orderings on Σ (see Dershowitz [7]).

The relative strength and the ordinals associated with these orderings have been studied by Okada and Dershowitz [37, 9]. For instance, given a strict ordering $\prec$ on a finite set Σ of n elements, then T_Σ under $\prec_{rpo}$ is order-isomorphic to $\varphi_n(0)$, the first n-critical

ordinal.[7] In particular, there is a very natural representation of the ordinals less than ϵ_0 in terms of nested multisets of natural numbers. It is even possible to define an *rpo* whose order-type is Γ_0 (see Dershowitz [7]), if we allow terms to serve as labels.[8]

Okada has showed that it is possible to combine the multiset and lexicographic ordering to obtain term orderings subsuming both the *rpo* and *lpo* ordering, and also obtain a system of notations for the ordinals less than Γ_0 (see Okada [37], and Dershowitz and Okada [9]). Such systems are inspired by some earlier work of Ackermann [1], and we briefly describe one of them.

Let C be a set of constants, and F a set of function symbols (we are not assuming that symbols in F have a fixed arity).

Definition 11.29 For any $n > 0$, the set $A_n(F, C)$ of *generalized Ackermann terms* is defined inductively as follows:

(1) $c \in A_n(F, C)$ whenever $c \in C$.

(2) $f(t_1, \ldots, t_n) \in A_n(F, C)$ whenever $f \in F$ and $t_1, \ldots, t_n \in A_n(F, C)$.

The terms defined by (1) and (2) are called *connected terms*.

(3) $t_1 \# \cdots \# t_m \in A_n(F, C)$, whenever $t_1, \ldots, t_m$ are connected terms in $A_n(F, C)$ ($m \geq 2$).[9]

Given a set $\Sigma = C \cup F$ of labels, note that the set of trees T_Σ can be viewed as a subset of $A_1(F, C)$, using the following representation function:

$rep(c) = c$, when $c \in C$, and

$rep(f(t_1, \ldots, t_m)) = f(rep(t_1) \# \cdots \# rep(t_n))$.

Given a preorder $\preceq$ on $C \cup F$, we define a preorder $\preceq_{ack}$ on $A_n(F, C)$ as follows.

Definition 11.30 The *Ackermann ordering* $\preceq_{ack}$ on $A_n(F, C)$ is defined below. Actually, we give a simultaneous recursively definition of $\succeq_{ack}$, $\succ_{ack}$, and $\approx_{ack}$, where $s \succ_{ack} t$ iff $s \succeq_{ack} t$ and $s \not\preceq_{ack} t$, and $s \approx_{ack} t$ iff $s \succeq_{ack} t$ and $s \preceq_{ack} t$.

(1) If $s, t \in C$, then $s \succeq_{ack} t$ iff $s \succeq t$. If $s \in C$ and $t \notin C$, then $t \succ_{ack} s$ (and $t \not\preceq_{ack} s$).

(2) Let $s = f(s_1, \ldots, s_n)$ and $t = g(t_1, \ldots, t_n)$. Then, $s \succeq_{ack} t$ iff one of the conditions below holds:

---

[7] In this case, Σ is not a ranked alphabet. We allow the symbols in Σ to have varying (finite) ranks.

[8] These terms are formed using a single symbol $\star$ that can assume any finite rank.

[9] Compared to the definition in Dershowitz and Okada [9], we require that $t_1, \ldots, t_m$ are connected terms. This seems cleaner and does not seem to cause any loss of generality.

(i) $f \approx g$, $s_1 \approx_{ack} t_1, \ldots, s_{i-1} \approx_{ack} t_{i-1}, s_i \succeq_{ack} t_i$, and $s \succ_{ack} t_{i+1}, \ldots, s \succ_{ack} t_n$, for some i, $1 \leq i \leq n$; or

(ii) $f \succ g$ and $f(s_1, \ldots, s_n) \succ_{ack} t_i$ for *all* i, $1 \leq i \leq n$; or

(iii) $s_i \succeq_{ack} g(t_1, \ldots, t_n)$ for *some* i, $1 \leq i \leq n$.

(3) Let $s = s_1 \# \cdots \# s_m$ (or $s = s_1$) and $t = t_1 \# \cdots \# t_p$ (or $t = t_1$). Then, $s \succeq_{ack} t$ iff

$$\{s_1, \ldots, s_m\} \succeq_{ack}^{mult} \{t_1, \ldots, t_p\},$$

where $\succeq_{ack}^{mult}$ is the multiset extension of $\succeq_{ack}$.

The following results are stated in Okada [37], and Dershowitz and Okada [9].

Theorem 11.31 (1) If the strict order $\succ$ is well-founded on $C \cup F$, then $\succ_{ack}$ is well-founded on $A_n(F, C)$.

(2) The multiset extension of *rpo* is identical to $\succ_{ack}$ on $A_1(F, C)$.

Proof. The proof of (1) uses the techniques used in theorem 11.22 (Kruskal's theorem). The proof of (2) is straightforward. □

Equivalently, part (2) of theorem 11.31 says that the restriction of $\succeq_{ack}$ to connected terms in $A_1(F, C)$ is identical to *rpo* (we use the representation of terms given by the function *rep* described earlier).

Finally, as noted by Okada, $\langle A_2(\{\psi\}, \{0\}), \preceq_{ack} \rangle$ provides a system of notations for the ordinals less than Γ_0. This is easily seen using theorem 8.12. To show that $\preceq_{ack}$ corresponds to the ordering on the ordinals less than Γ_0, we use lemma 8.11 and lemma 8.10. We can even define a bijection *ord* bewteen the equivalence classes of $A_2(\{\psi\}, \{0\})$ modulo $\approx_{ack}$ and the set of ordinals less than Γ_0 as follows:

$ord(\psi(s, t)) = \psi(ord(s), ord(t))$,

$ord(s_1 \# \cdots \# s_m) = \alpha_1 + \cdots + \alpha_m$,

where $\alpha_1 \geq \ldots \geq \alpha_m$ is the sequence obtained by ordering $\{ord(s_1), \ldots, ord(s_m)\}$ in nonincreasing order.

12 A Glimpse at Hierarchies of Fast and Slow Growing Functions

In this section, we discuss briefly some hierarchies of functions that play an important role in logic because they provide natural classifications of recursive functions according to their computational complexity. It is appropriate to discuss these classes of functions now,

because we have sufficient background about constructive ordinal notations at our disposal. When restricted to the ordinals less than ϵ_0, these hierarchies provide natural rate-of-growth and complexity classifications of the recursive functions which are *provably total* in Peano's arithmetic. In particular, for two of these hierarchies, F_{ϵ_0} and H_{ϵ_0} dominate every such function (for all but finitely many arguments). Thus, the statement "F_{ϵ_0} is total recursive" is true, but not provable in Peano's arithmetic. The relationship with Kruskal's theorem is that the function Fr mentioned in the discussion following theorem 5.2 dominates F_{ϵ_0} (for all but finitely many arguments). In fact, Fr has the rate of growth of a function F_α where α is considerably larger that Γ_0! The results of this section are presented in Cichon and Wainer [4], and Wainer [54], and the reader is referred to these papers for further details.

For ease of understanding, we begin by defining hierarchies indexed by the natural numbers. There are three classes of hierarchies.

1. *Outer iteration hierarchies.*

Let $g: \mathbf{N} \to \mathbf{N}$ be a given function. The hierarchy $(g_m)_{m\in\mathbf{N}}$ is defined as follows: For all $n \in \mathbf{N}$,

$$g_0(n) = 0,$$
$$g_{m+1}(n) = g(g_m(n)).$$

The prime example of this kind of hierarchy is the *slow-growing hierarchy* $(G_m)_{m\in\mathbf{N}}$ based on the successor function $g(n) = n + 1$. This hierarchy is actually rather dull when the G_m are indexed by finite ordinals, since $G_m(n) = m$ for all $n \in \mathbf{N}$, but it is much more interesting when the index is an infinite ordinal.

2. *Inner iteration hierarchies.*

Again, let $g: \mathbf{N} \to \mathbf{N}$ be a given function. The hierarchy $(h_m)_{m\in\mathbf{N}}$ is defined as follows: For all $n \in \mathbf{N}$,

$$h_0(n) = n,$$
$$h_{m+1}(n) = h_m(g(n)).$$

The prime example of this kind of hierarchy is the *Hardy hierarchy* $(H_m)_{m\in\mathbf{N}}$ based on the successor function $g(n) = n+1$. This hierarchy is also rather dull when the H_m are indexed by finite ordinals, since $H_m(n) = n+m$ for all $n \in \mathbf{N}$, but it is much more interesting when the index is an infinite ordinal.

3. *Fast iteration hierarchies.*

Let $g: \mathbf{N} \to \mathbf{N}$ be a given increasing function. The hierarchy $(f_m)_{m \in \mathbf{N}}$ is defined as follows: For all $n \in \mathbf{N}$,

$$f_0(n) = g(n),$$
$$f_{m+1}(n) = f_m^n(n),$$

where $f_m^n(x) = f_m(f_m(\ldots(f_m(x))\ldots))$, the nth iterate of f_m applied to x. The prime example of this kind of hierarchy is the *Grzegorczyk hierarchy* $(F_m)_{m \in \mathbf{N}}$ based on the successor function $g(n) = n + 1$. This hierarchy is not dull even when the F_m are indexed by finite ordinals. Indeed, $F_1(n) = 2n$, $F_2(n) = 2^n n$, and

$$2^{2^{\cdot^{\cdot^{\cdot^{2^n}}}}\Big\}n} < F_3(n).$$

In order to get functions growing even faster than those obtained so far, we extend these hierarchies to infinite ordinals. The trick is to diagonalize at limit ordinals. However, this presuposes that for each limit ordinal α under consideration, we already have a particular predefined increasing sequence $\alpha[0], \alpha[1], \ldots, \alpha[n], \ldots$, such that $\alpha = \bigsqcup_{n \in \mathbf{N}} \alpha[n]$, a so-called *fundamental sequence*. The point of ordinal notations is that they allow the definition of standard fundamental sequences. This is particularly simple for the ordinals less than ϵ_0, where we can use the Cantor normal form.

For every limit ordinal $\delta < \epsilon_0$, if $\delta = \alpha + \beta$, then $\delta[n] = \alpha + \beta[n]$, if $\delta = \omega^{\alpha+1}$, then $\delta[n] = \omega^\alpha n$ (i.e. $\omega^\alpha + \cdots + \omega^\alpha$ n times), and when $\delta = \omega^\alpha$ for a limit ordinal α, then $\delta[n] = \omega^{\alpha[n]}$. For ϵ_0 itself, we choose $\epsilon_0[0] = 0$, and $\epsilon_0[n+1] = \omega^{\epsilon_0[n]}$.

Fundamental sequences can also be assigned to certain classes of limit ordinals larger than ϵ_0, but this becomes much more complicated. In particular, this can be done for limit ordinals less than Γ_0, using the normal form representation given in theorem 8.2.

Assuming that fundamental sequences have been defined for all limit ordinals in a given subclass $\mathcal{I}$ of $\mathcal{O}$, we extend the definition of the hierarchies as follows.

Definition 12.1 *Outer iteration hierarchies.*

Let $g: \mathbf{N} \to \mathbf{N}$ be a given function. The hierarchy $(g_\alpha)_{\alpha \in \mathcal{I}}$ is defined as follows: For all $n \in \mathbf{N}$,

$$g_0(n) = 0,$$
$$g_{\alpha+1}(n) = g(g_\alpha(n)),$$
$$g_\alpha(n) = g_{\alpha[n]}(n),$$

where in the last case, α is a limit ordinal. The prime example of this kind of hierarchy is the *slow-growing hierarchy* $(G_\alpha)_{\alpha\in\mathcal{I}}$ based on the successor function $g(n) = n + 1$. This time, we can show that for any n, $g_\alpha(n) = g^{G_\alpha(n)}(0)$, and $G_{\alpha+\beta}(n) = G_\alpha(n) + G_\beta(n)$, from which it follows that $G_{\omega^\alpha}(n) = n^{G_\alpha(n)}$. This means that if α is represented in Cantor normal form, then $G_\alpha(n)$ is the result of replacing ω by n troughout the Cantor normal form! Thus, we have

$$G_{\epsilon_0[n]}(n) = n^{n^{\cdot^{\cdot^{\cdot^{n^n}}}}\Big\}n-1}.$$

Definition 12.2 *Inner iteration hierarchies.*

Again, let $g\colon \mathbf{N} \to \mathbf{N}$ be a given function. The hierarchy $(h_\alpha)_{\alpha\in\mathcal{I}}$ is defined as follows: For all $n \in \mathbf{N}$,

$$\begin{aligned} h_0(n) &= n, \\ h_{\alpha+1}(n) &= h_\alpha(g(n)), \\ h_\alpha(n) &= h_{\alpha[n]}(n), \end{aligned}$$

where in the last case, α is a limit ordinal. The prime example of this kind of hierarchy is the *Hardy hierarchy* $(H_\alpha)_{\alpha\in\mathcal{I}}$ based on the successor function $g(n) = n + 1$ (Hardy [20]). It is easy to show that $h_{\alpha+\beta}(n) = h_\alpha(h_\beta(n))$, and so $h_{\omega^{\alpha+1}}(n) = h^n_{\omega^\alpha}(n)$.

Definition 12.3 *Fast iteration hierarchies.*

Let $g\colon \mathbf{N} \to \mathbf{N}$ be a given increasing function. The hierarchy $(f_\alpha)_{\alpha\in\mathcal{I}}$ is defined as follows: For all $n \in \mathbf{N}$,

$$\begin{aligned} f_0(n) &= g(n), \\ f_{\alpha+1}(n) &= f^n_\alpha(n), \\ f_\alpha(n) &= f_{\alpha[n]}(n), \end{aligned}$$

where $f^n_\alpha(x) = f_\alpha(f_\alpha(\ldots(f_\alpha(x))\ldots))$, the nth iterate of f_α applied to x, and in the last case, α is a limit ordinal.

The prime example of this kind of hierarchy is the extended *Grzegorczyk hierarchy* $(F_\alpha)_{\alpha\in\mathcal{I}}$ based on the successor function $g(n) = n + 1$. It is interesting to note that Ackermann's function has rate of growth roughly equivalent to that of F_ω.

It is not difficult to show that $f_\alpha(n) = h_{\omega^\alpha}(n)$. Thus, even though the fast-growing hierarchy seems to grow faster than the inner iteration hierarchy, the h-hierarchy actually "catches up" with the f-hierarchy at ϵ_0, in the sense that

$$f_{\epsilon_0}(n-1) \le h_{\epsilon_0}(n) \le f_{\epsilon_0}(n+1).$$

Given two functions $f, g: \mathbf{N} \to \mathbf{N}$, we say that g *majorizes* f *(or that* g *dominates* f*)* iff there is some $k \in \mathbf{N}$ such that $g(n) > f(n)$ for all $n \geq k$. It is shown in Buchholz and Wainer [3] that F_β majorizes F_α and that H_β majorizes H_α if $\beta > \alpha$. This property can also be shown for the slow-growing hierarchy. Buchholz and Wainer [3] also show that every recursive function provably total in Peano's arithmetic is majorized by some $F_{\alpha+1}$ in the fast-growing hierarchy up to ϵ_0, and that every F_α for $\alpha < \epsilon_0$ is recursive and provably total in *PA*. It follows that F_{ϵ_0} is recursive, but *not* provably total in *PA*. Going back to the function Fr associated with Friedman miniature version of Kruskal's theorem (theorem 5.2), Friedman has shown that Fr majorizes F_{Γ_0}, and in fact, Fr has the rate of growth of a function F_α where α is considerably larger that Γ_0!

We noted that the h-hierarchy catches up with the f-hierarchy at ϵ_0. It is natural to ask whether the slow-growing hierarchy catches up with the fast-growing hierarchy. At first glance, one might be skeptical that this could happen. But large ordinals are tricky objects, and in fact there is an ordinal α such that the slow-growing hierarchy catches up with the fast-growing hierachy.

Theorem 12.4 (Girard) There is an ordinal α such that G_α and F_α have the same rate of growth, in the sense that

$$G_\alpha(n) < F_\alpha(n) < G_\alpha(an + b),$$

for some simple linear function $an + b$. □

This remarkable result was first proved by Girard [17]. The ordinal α for which G_α and F_α have the same rate of growth is nonother than *Howard's ordinal*, another important ordinal occurring in proof theory. Unfortunately, we are not equipped to describe it, even with the apparatus of the normal functions $\varphi(\alpha, \beta)$. Howard's ordinal is greater than Γ_0, and it is denoted by $\varphi_{\epsilon_{\Omega+1}+1}(0)$, where Ω is the least uncountable ordinal, and $\epsilon_{\Omega+1}$ is the least ϵ-number after Ω (so $\epsilon_{\Omega+1} = \Omega^{\Omega^{\Omega^{\cdots}}}$). Alternate proofs of this result are given in Cichon and Wainer [4], and Wainer [54] (among others). A fairly simple description of Howard's ordinal is given in Pohlers [41].

Before closing this section, we cannot resist mentioning Goodstein sequences [18], another nice illustration of the representation of ordinals less than ϵ_0 in Cantor normal form.

Let n be any fixed natural number, and consider any natural number a such that

$$a < (n+1)^{(n+1)^{\cdots^{(n+1)}}\}(n+1)}.$$

We express a in *complete base* $n+1$ by first writing $a = m_0 + m_1(n+1) + \ldots + m_k(n+1)^{a_k}$, where $m_i \leq n$, and $a_i < a_{i+1}$, and then recursively writing each a_i in complete base $n+1$, until all the exponents are $\leq n$. Given a, denote by $rep(a, n+1)$ its associated representation in complete base $n+1$. Given a number a and its representation $rep(a, n+1)$, we denote by $shiftrep(a, n+1)$ the result of replacing $n+1$ by $n+2$ throughout the representation $rep(a, n+1)$, and by $|shiftrep(a, n+1)|$ the numerical value of this new term.

Definition 12.5 The *Goodstein sequence starting with* $a \geq 0$ is defined as follows. Choose n as the least number such that

$$a < (n+1)^{(n+1)^{\cdot^{\cdot^{\cdot^{(n+1)}}}}\}(n+1)}.$$

Set $a_0 = a - 1$, and $a_{k+1} = |shiftrep(a_k, n+k+1)| - 1$.

In the above definition, $a - b$ is the usual difference between a and b when $a \geq b$, and it is equal to 0 otherwise. Thus, we obtain a_{k+1} from a_k by changing $n+k+1$ to $n+k+2$ in the representation $rep(a_k, n+k+1)$ of a_k and subtracting 1 from this new value.

Theorem 12.6 (Goodstein, Kirby and Paris) Every Goodstein sequence terminates, that is, there is some k such that $a_k = 0$. Furthermore, the function *Good* such that $Good(a) =$ the least k such that $a_k = 0$ is recursive, but it majorizes the function H_{ϵ_0} from the Hardy Hierarchy.

Proof. The proof that every Goodstein sequence terminates is not that difficult. The trick is to associate to each a_k an ordinal $\alpha_k < \epsilon_0$ obtained by replacing $n+k+1$ by ω throughout $rep(a_k, n+k+1)$. Then, it is easy to see that $\alpha_{k+1} < \alpha_k$, and thus, the sequence a_k reaches 0 for some k. The second part of the theorem is due to Kirby and Paris [26]. Another relatively simple proof appears in Buchholz and Wainer [3]. □

Since H_{ϵ_0} is not provably recursive in *PA*, Goodstein's theorem is a statement that is true but not provable in *PA*.

Readers interested in combinatorial independence results are advised to consult the beautiful book on Ramsey theory, by Graham, Rothschild, and Spencer [19].

13 Constructive Proofs of Higman's Lemma

If one looks closely at the proof of Higman's lemma (lemma 3.2), one notices that the proof is not constructive for two reasons:

(1) The proof proceeds by contradiction, and thus it is not intuitionistic.

(2) The definition of a minimal bad sequence is heavily impredicative, as it involves universal quantification over **all** bad sequences.

Thus, it is natural, and as it turns out, quite challenging, to ask whether it is possible to give a constructive (and predicative) proof of Higman's lemma.

In a remarkable (and short) paper, Friedman [15] introduces a new and simple technique, *the A-translation*, which enables him to give simple proofs of the fact that first-order classical Peano arithmetic and classical higher-order arithmetic are conservative over their respective intuitionistic version over Π^0_2-sentences. His technique also yields closure under Markov's rule for several intuitionistic versions of arithmetic (if $\neg\neg\exists x\varphi$ is provable, then $\exists x\varphi$ is also provable, where x is a numeric variable, and φ is a primitive recursive relation). Using Friedman's A-translation technique, it follows that there is an intuitionistic impredicative proof of Higman's lemma. However, it would still be interesting to see whether a constructive (predicative) proof can be extracted *directly* from the classical proof, and Gabriel Stolzenberg was among the first researchers to propose this challenge, and eventually solve it. It turns out that (at least) two constructive (predicative) proofs of a constructive version of Higman's lemma have been given independently by Richman and Stolzenberg [45], and Murthy and Russell [35]. Steve Simpson has proven a related result for the Hilbert's basis theorem [49], and his proof technique seems related to some of the techniques of Richman and Stolzenberg. The significance of having a constructive proof is that one gets an algorithm which, given a constructively (and finitely presented) infinite sequence, yields the lefmost pair of embedded strings. Murthy and Russell [35] do extract such an algorithm using the NuPRL proof development system. The next challenge is to find a constructive proof of Kruskal's theorem.

Acknowledgment: I wish to thank Robert Constable, Thierry Coquand, Nachum Dershowitz, Jean-Yves Girard, Pierre Lescanne, Anil Nerode, Mitsu Okada, Wayne Snyder, Rick Statman, and Gabriel Stolzenberg, for helpful comments and for pointing out related work.

14 References

[1] Ackermann, W. Konstruktiver Aubfau eines Abschnitts der zweiten Cantorschen Zalhenklasse. *Math. Zeit.* 53, 403-413 (1951).

[2] Bachmair, L. *Canonical Equational Proofs.* John Wiley and Sons, New York (1990).

[3] Buchholz, W., and Wainer, S.S. Provably Computable Functions and the Fast Growing Hierarchy. *Logic and Combinatorics*, edited by S. Simpson, Comtemporary Math-

ematics, Vol. 65, AMS (1987), 179-198.

[4] Cichon, E.A., and Wainer, S.S. The Slow-Growing and the Grzegorczyk Hierarchies. *J. of Symbolic Logic* 48(2) (1983), 399-408.

[5] Crossley, J.N., and Bridge Kister, J. Natural Well-Orderings. *Arch. math. Logik* 26 (1986/1987), 57-76.

[6] DeJongh, D.H.J., and Parikh, R. Well partial orderings and hierarchies. *Indagationes Mathematicae* 14 (1977), 195-207.

[7] Dershowitz, N. Termination of Rewriting. *J. Symbolic Computation* (3), 1-2 (1987), 69-116.

[8] Dershowitz, N. Orderings for Term-Rewriting Systems. *TCS* 17(3) (1982), 279-301.

[9] Dershowitz, N., and Okada, M. Proof-theoretic techniques for term rewriting theory. *3rd Annual Symposium on Logic In Computer Science*, IEEE Computer Society, Edinburgh, Scotland, July 1988, 104-111.

[10] Dershowitz, N., and Manna, Z. Proving termination with multiset orderings. *Communications of the ACM* 22, 465-476 (1979).

[11] Dershowitz, N. Completion and its Applications. In *Resolution of Equations in Algebraic Structures*, Vol. 2, Aït-Kaci and Nivat, editors, Academic Press, 31-85 (1989).

[12] Dickson, L.E. Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors. *Am. J. Math* 35, (1913), 413-426.

[13] Feferman, S. Systems of Predicative Analysis. *J. of Symbolic Logic* 29 (1964), 1-30.

[14] Feferman, S. Proof Theory: A Personnal Report. In [52], 447-485.

[15] Friedman, H. Classically and intuitionistically provably recursive functions. *Higher set theory* (G.H. Müller and Dana S. Scott, editors), Lecture Notes in Mathematics, Vol. 699, Springer-Verlag, Berlin (1978), 21-28.

[16] Friedman, H., McAloon, K., and Simpson, S. A finite combinatorial principle which is equivalent to the 1-consistency of predicative analysis. *Logic Symposion I (Patras, Greece, 1980)*, G. Metakides, editor, North-Holland, Amsterdam, (1982), 197-230.

[17] Girard, J.Y. Π^1_2-logic. *Annals of Mathematical Logic* 21 (1981), 75-219.

[18] Goodstein, R.L. On the restricted ordinal theorem. *J. of Symbolic Logic* 9 (1944), 33-41.

[19] Graham, R.L., Rothschild, B.L., and Spencer, J.H. *Ramsey Theory*, John Wiley & Sons, Inc., 2nd edition, pp. 196 (1990).

[20] Hardy, G.H. A theorem concerning the infinite cardinal numbers. *Quarterly J. Math.* 35 (1904), 87-94.

[21] Harrington, L.A. et al. *Harvey Friedman's Research on the Foundations of Mathematics.* Harrington, Morley, Ščedrov, and Simpson, Editors, North-Holland (1985).

[22] Higman, G. Ordering by divisibility in abstract algebras. *Proc. London Math. Soc.* (3), 2 (1952), 326-336.

[23] Janet, M. Sur les systèmes d'équations aux dérivées partielles. *J. de Mathématiques* III(8) (1920).

[24] Jouannaud, J.P., and Lescanne, P. On multiset orderings. *Information Processing Letters* 15(2), 57-63 (1982).

[25] Kaplanski, I. Ph.D. thesis, 1941.

[26] Kirby L., and Paris, J. Accessible independence results from Peano arithmetic. *Bull. London Math. Soc.* 14 (1982), 285-293.

[27] Knuth, D.E. and Bendix, P.B., "Simple Word Problems in Univeral Algebras," in *Computational Problems in Abstract Algebra*, Leech, J., ed., Pergamon Press (1970).

[28] Kruskal, J.B. Well-quasi-ordering, the tree theorem, and Vázsonyi's conjecture. *Trans. American Math. Soc.* 95 (1960), 210-225.

[29] Lescanne, P. Some properties of decomposition ordering. A simplification ordering to prove termination of rewriting systems. *RAIRO Informatique Théorique* 16(4), 331-347 (1982).

[30] Lescanne, P. Well rewrite orderings. Extended Abstract, Centre de Recherche en Informatique de Nancy, France (1989).

[31] Lescanne, P. Well quasi orderings in a paper by Maurice Janet. *Bulletin of the EATCS*, No. 39, (1989), 185-188.

[32] Lescanne, P. On the recursive decomposition ordering with lexicographical status and other related orderings. To appear in *Journal of Automated Reasoning* (1990).

[33] Levy, J.J. "Kruskaleries et Dershowitzeries", unpublished notes (1981).

[34] Miller, L.W. Normal functions and constructive ordinal notations. *J. of Symbolic Logic* 41(2) (1976), 439-459.

[35] Murthy, C.R., and Russell, J.R. A constructive proof of Higman's lemma. *5th Annual Symposium on Logic In Computer Science*, IEEE Computer Society, Philadelphia, PA, 257-267, June 4-7, 1990.

[36] Nash-Williams, C. St. J. A. On well-quasi-ordering finite trees. *Proc. Cambridge Phil. Soc.* 59 (1963), 833-835.

[37] Okada, M. Ackermann's ordering and its relationship with ordering systems of term rewriting theory. *Proceedings of the 24th Allerton Conference on Communication, Control, and Computing*, Monticello, ILL (1986).

[38] Okada, M., and Takeuti. G. On the theory of quasi ordinal diagrams. *Logic and Combinatorics*, edited by S. Simpson, Comtemporary Mathematics, Vol. 65, AMS (1987), 295-307.

[39] Okada, M. Kruskal-type theorems on labeled finite trees in term-rewriting theory, graph theory, and proof theory. Manuscript (1987).

[40] Okada, M. Quasi-ordinal diagrams and Kruskal-type theorems on labeled finite trees. Manuscript (1987).

[41] Pohlers, W. *Proof Theory, an Introduction.* Lecture Notes in Mathematics No. 1407, Springer Verlag (1989).

[42] Pohlers, W. Proof theory and ordinal analysis. Preprint, MSRI, Berkeley, California (1989)

[43] Puel, L. Bon préordres sur les arbres associés à des ensembles inévitables et preuves de terminaison de systèmes de réécriture. Thèse d'Etat, (1987), Université de Paris VII.

[44] Puel, L. Using unavoidable sets of trees to generalize Kruskal's theorem. Technical Report 86-4, Laboratoire d'Informatique de l'Ecole Normale Supérieure, Paris, France (1986).

[45] Richman, F., and Stolzenberg, G. Well quasi-ordered sets. Technical report submitted for publication, Northeastern University, Boston MA, and Harvard University, Cambridge, MA, April 1990.

[46] Schütte, K. *Proof Theory.* Springer-Verlag (1977).

[47] Simpson, S.G. Nonprovability of certain combinatorial properties of finite trees. In [21], 87-117.

[48] Simpson, S.G. Which set existence axioms are needed to prove the Cauchy/Peano theorem for ordinary differential equations? *J. of Symbolic Logic* 49(3) (1984), 783-802.

[49] Simpson, S.G. Ordinal numbers and the Hilbert basis theorem. *Journal of Symbolic Logic* 53 (1988), 961-964.

[50] Smoryński, C. "Big" News From Archimedes to Friedman. In [21], 353-366.

[51] Smoryński, C. The Varieties of Arboreal Experience. In [21], 381-397.

[52] Takeuti, G. *Proof Theory.* Studies in Logic, Vol. 81, North-Holland, Amsterdam, Second Edition (1987).

[53] Veblen, O. Continuous increasing functions of finite and transfinite ordinals. *Transactions of the American Mathematical Society*, Vol. 9 (1908), 280-292.

[54] Wainer, S.S. Slow Growing Versus Fast Growing. *J. of Symbolic Logic* 54(2) (1989), 608-614.